256509

STIC-ILL

From: Enewold, Jeanine
Sent: Thursday, July 08, 1999 8:44 AM
To: STIC-ILL
Subject: Art- 09137822

Hi-
Could you please find and copy these articles.

1. Bioorg Med Chem 1996 Jan;4(1):5-23
"Peptide nucleic acids (PNA): synthesis, properties and potential applications."
Hyrup B, Nielsen PE

2. Science 1995 Dec 15;270(5243):1838-41
"A nucleic acid triple helix formed by a peptide nucleic acid-DNA complex."
Betts L, Josey JA, Veal JM, Jordan SR

3. Annu Rev Biochem 1995;64:65-95
"Triplex DNA structures."
Frank-Kamenetskii MD, Mirkin SM

4. J Biol Chem 1995 Jun 9;270(23):14068-71
"An unusually stable purine(purine-pyrimidine) short triplex. The third strand stabilizes double-stranded DNA."
Svinarchuk F, Paoletti J, Malvy C

Thank you

Jeanine Enewold
Art Unit: 1655
Room 12-D11
306-5817

PTO
7/8
WOS

Article Title	TRIPLEX DNA STRUCTURES
Article Identifier	006641549510003Z
Authors	Frank-Kamenetskii_M_D Mirkin_S_M
Journal Title	Annual Review of Biochemistry
Publisher	Annual Reviews
ISSN	0066-4154
Volume	64
Issue	1995
Supplement	0
Year of Publication	1995
Page Range	65-95
Number of Pages	31
User Name	Adonis
Cost Centre	Development
PCC	$8.00
Date and Time	Thursday, July 08, 1999 7:35:02 PM

Annu. Rev. Biochem. 1995. 64:65–95

TRIPLEX DNA STRUCTURES

Maxim D. Frank-Kamenetskii
Center for Advanced Biotechnology and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

Sergei M. Mirkin
Department of Genetics, University of Illinois at Chicago, Chicago, Illinois 60612

KEY WORDS: DNA triplex, H-DNA, gene-drugs, triplex-forming oligonucleotides, PNA

CONTENTS

ABSTRACT

A DNA triplex is formed when pyrimidine or purine bases occupy the major groove of the DNA double Helix forming Hoogsteen pairs with purines of the Watson-Crick basepairs. Intermolecular triplexes are formed between triplex forming oligonucleotides (TFO) and target sequences on duplex DNA. Intramolecular triplexes are the major elements of H-DNAs, unusual DNA structures, which are formed in homopurine-homopyrimidine regions of supercoiled DNAs. TFOs are promising gene-drugs, which can be used in an anti-gene strategy, that attempt to modulate gene activity in vivo. Numerous chemical modifications of TFO are known. In peptide nucleic acid (PNA), the sugar-phosphate backbone is replaced with a protein-like backbone. PNAs form

0066-4154/95/0701-0065$05.00

P-loops while interacting with duplex DNA forming triplex with one of DNA strands leaving the other strand displaced. Very unusual recombination or parallel triplexes, or R-DNA, have been assumed to form under RecA protein in the course of homologous recombination.

PERSPECTIVES AND SUMMARY

Since the pioneering work of Felsenfeld, Davies, & Rich (1), double-stranded polynucleotides containing purines in one strand and pyrimidines in the other strand [such as poly(A)/poly(U), poly(dA)/poly(dT), or poly(dAG)/ poly(dCT)] have been known to be able to undergo a stoichiometric transition forming a triple-stranded structure containing one polypurine and two polypyrimidine strands (2–4). Early on, it was assumed that the third strand was located in the major groove and associated with the duplex via non-Watson-Crick interactions now known as Hoogsteen pairing. Triple helices consisting of one pyrimidine and two purine strands were also proposed (5, 6). However, notwithstanding the fact that single-base triads in tRNA structures were well-documented (reviewed in 7), triple-helical DNA escaped wide attention before the mid-1980s.

The considerable modern interest in DNA triplexes arose due to two partially independent developments. First, homopurine-homopyrimidine stretches in supercoiled plasmids were found to adopt an unusual DNA structure, *called* H-DNA, which includes a triplex as the major structural element (8, 9). Secondly, several groups demonstrated that homopyrimidine and some purine-rich oligonucleotides can form stable and sequence-specific complexes with corresponding homopurine-homopyrimidine sites on duplex DNA (10–12). These complexes were shown to be triplex structures rather than D-loops, where the oligonucleotide invades the double helix and displaces one strand. A characteristic feature of all these triplexes is that the two chemically homologous strands (both pyrimidine or both purine) are antiparallel. These findings led to explosive growth in triplex studies.

During the study of intermolecular triplexes, it became clear that triplex-forming oligonucleotides (TFOs) might be universal drugs that exhibit sequence-specific recognition of duplex DNA. This is an exciting possibility because, in contrast to other DNA-binding drugs, the recognition principle of TFOs is very simple: Hoogsteen pairing rules between a purine strand of the DNA duplex and the TFO bases. However, this mode of recognition is limited in that homopurine-homopyrimidine sites are preferentially recognized. Though significant efforts have been directed toward overcoming this limitation, the problem is still unsolved in general. Nevertheless, the high specificity of TFO-DNA recognition has led to the development of an "antigene" strategy, the goal of which is to modulate gene activity in vivo using TFOs (reviewed in 13).

Although numerous obstacles must be overcome to reach the goal, none are likely to be fatal for the strategy. Even if DNA TFOs proved to be unsuitable

as gene-drugs, there are already many synthetic analogs that also exhibit triplex-type recognition. Among them are oligonucleotides with non-natural bases capable of binding the duplex more strongly than can natural TFOs. Another promising modification replaces the sugar-phosphate backbone of ordinary TFO with an uncharged peptidelike backbone, called a peptide nucleic acid (PNA) (reviewed in 14). Homopyrimidine PNAs form remarkably strong and sequence-specific complexes with the DNA duplex via an unusual strand-displacement reaction: Two PNA molecules form a triplex with one of the DNA strands, leaving the other DNA strand displaced (a "P-loop") (15, 16).

The ease and sequence specificity with which duplex DNA and TFOs formed triplexes seemed to support the idea (17) that the homology search preceding homologous recombination might occur via a triplex between a single DNA strand and the DNA duplex without recourse to strand separation in the duplex. However, these proposed recombination triplexes are dramatically different from the orthodox triplexes observed experimentally. First, the recombination triplexes must be formed for arbitrary sequences and, second, the two identical strands in this triplex are parallel rather than antiparallel. Some data supported the existence of a special class of recombination triplexes, at least within the complex among duplex DNA, RecA protein, and single-stranded DNA (reviewed in Ref. 18), called R-DNA. A stereochemical model of R-DNA was published (19). However, the structure of the recombination intermediate is far from being understood, and some recent data strongly favor the traditional model of homology search via local strand separation of the duplex and D-loop formation mediated by RecA protein.

Intramolecular triplexes (H-DNA) are formed in vitro under superhelical stress in homopurine-homopyrimidine mirror repeats. The average negative supercoiling in the cell is not sufficient to induce H-DNA formation in most cases. However, H-DNA can be detected in vivo in association with an increase of DNA supercoiling driven by transcription or other factors (reviewed in 20). H-DNA may even be formed without DNA supercoiling during in vitro DNA synthesis. Peculiarly, this DNA polymerase–driven formation of H-DNA efficiently prevents further DNA synthesis (21, 22). There are preliminary indications that H-DNA may also terminate DNA replication in vivo (23). More work is required, however, to elucidate the role of H-DNA in biological systems.

STRUCTURE, STABILITY, AND SPECIFICITY OF DNA TRIPLEXES

Triplex Menagerie

Since the original discovery of oligoribonucleotide-formed triplexes, numerous studies have shown that the structure of triplexes may vary substantially. First, it was shown that triplexes may consist of two pyrimidine and one purine strand

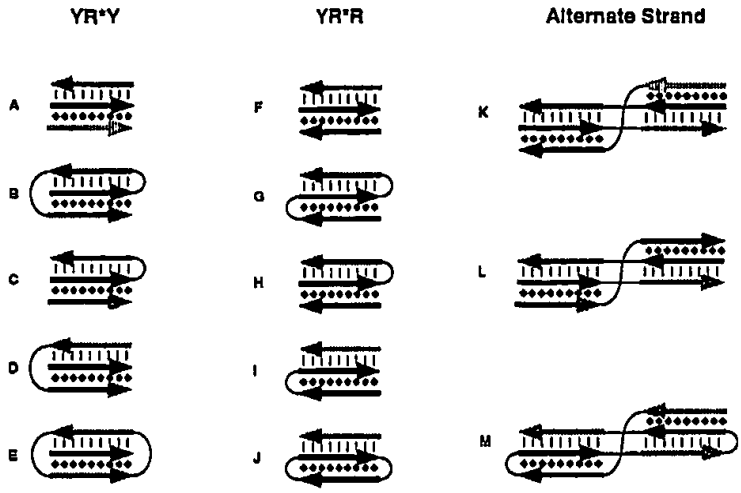

Figure 1 Triplex menagerie (see text for explanations). Solid lines, purine strands; stippled lines, pyrimidine strands; vertical lines, Watson-Crick hydrogen bonds; diamonds, Hoogsteen hydrogen bonds. Arrows indicate DNA chain polarity.

(YR*Y) or of two purine and one pyrimidine strand (YR*R). Second, triplexes can be built from RNA or DNA chains or their combinations. Third, triplexes can be formed within a single polymer molecule (intramolecular triplexes) or by different polynucleotides (intermolecular triplexes). Finally, for special DNA sequences consisting of clustered purines and pyrimidines in the same strand, triplex formation may occur by a strand-switch mechanism (alternate strand triplexes). Figure 1 summarizes numerous possible structures of triple-helical nucleic acids.

The building blocks of YR*Y triplexes are the canonical CG*C and TA*T triads shown in Figure 2. To form such triads, the third strand must be located in the major groove of the double helix that is forming Hoogsteen hydrogen bonds (24) with the purine strand of the duplex. The remarkable isomorphism of both canonical triads makes it possible to form a regular triple helix. This limits YR*Y triplexes to homopurine-homopyrimidine sequences in DNA. An important feature of the YR*Y triplexes is that formation of the CG*C triad requires the protonation of the N3 of cytosine in the third strand. Thus, such triplexes are favorable under acidic conditions (3, 4). By contrast, the YR*R triplexes usually do not require protonation (see below).

The mutual orientation of the chemically homologous strands in a triplex

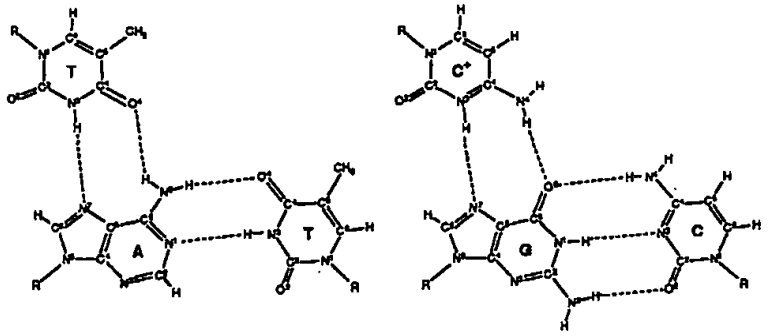

Figure 2 Canonical base triads of YR*Y triplexes: TA*T and CG*C$^+$.

(i.e. two pyrimidine strands in the YR*Y triplex or the two purine strands in the YR*R triplex), which a priori can be either parallel or antiparallel, is of paramount importance. The discovery of H- and *H-DNA (see below) indicated that both YR*Y and YR*R triplexes form as antiparallel structures (9, 25). A thorough investigation of intermolecular triplexes by different methods unambiguously demonstrated that both YR*Y and YR*R triplexes are stably formed only as antiparallel structures. The most direct data were obtained by cleaving target DNA with homopyrimidine or homopurine oligonucleotides attached to Fe•EDTA (11, 26). The observed pairing and orientation rules rigorously determine the sequence of the triplex-forming homopyrimidine strand.

YR*R triplexes are more versatile than YR*Y triplexes. Originally it was believed that they must be built from CG*G and TA*A triads (6, 25, 27). Later work showed, however, that TA*T triad may also be incorporated into the otherwise YR*R triplex. Moreover, the stability of triplexes consisting of alternating CG*G and TA*T triads is higher than that of triplexes built of CG*G and TA*A triads (26). Thus, the term YR*R triplex, though routinely used in literature, is misleading with regard to the chemical nature of the third strand. The corresponding triads are presented in Figure 3. One can see that they are not strictly isomorphous, as was the case for YR*Y triads. Another notable difference between two triplex types is that reverse Hoogsteen base-pairs are needed to form reasonable stacking interactions among CG*G, TA*A, and TA*T triads (26).

Whereas in YR*Y triplexes the sequence of the third strand is fully determined by the sequence of the duplex, the situation is different for YR*R triplexes. Here the third strand may consist of three bases, G, A, and T, where

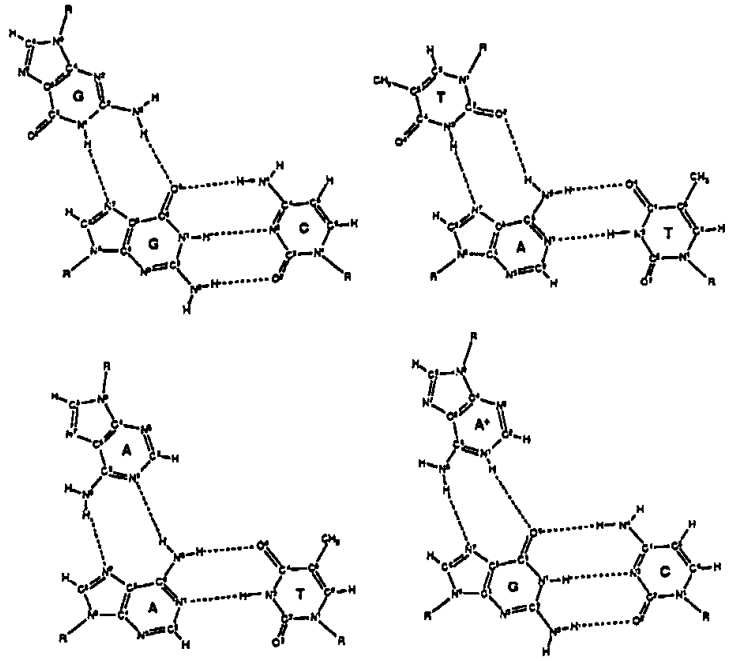

Figure 3 Base triads of YR*R triplexes: CG*G, TA*A, TA*T, and CG*A$^+$.

the guanines oppose guanines in the duplex, while adenines or thymines must oppose adenines of the duplex. A protonated CG*A$^+$ triad (Figure 3) forms so that A in the third strand may also oppose G in the duplex at acidic pH (28).

Another novel feature of YR*R triplexes is that their stability depends dramatically on the presence of bivalent metal cations (reviewed in 20). Unlike the case of YR*Y triplexes, where the requirement for H^+ ions has an obvious reason, the metal dependence of YR*R triplexes is an obscure function of the particular metal ion and the triplex sequence (29). Possible structural reasons for selectivity of bivalent cations in stabilization of YR*R triplexes are discussed in Ref. 30.

Despite these differences, the YR*R triplexes are similar to YR*Y triplexes in their most fundamental features: (*a*) The duplex involved in triplex formation must have a homopurine sequence in one strand, and (*b*) the orientation of the two chemically homologous strands is antiparallel.

One can easily imagine numerous "geometrical" ways to form a triplex, and those that have been studied experimentally are shown in Figure 1. The canonical intermolecular triplex consists of either three independent oligonucleotide chains (3, 4) (Figure 1*A,F*) or of a long DNA duplex carrying the homopurine-homopyrimidine insert and the corresponding oligonucleotide (10–12). In any case, triplex formation strongly depends on the oligonucleotide(s) concentration.

A single DNA chain may also fold into a triplex connected by two loops (Figure 1*B,G*). To comply with the sequence and polarity requirements for triplex formation, such a DNA strand must have a peculiar sequence: It contains a mirror repeat (homopyrimidine for YR*Y triplexes and homopurine for YR*R triplexes) flanked by a sequence complementary to one half of this repeat (31). Such DNA sequences fold into triplex configuration much more readily than do the corresponding intermolecular triplexes, because all triplex-forming segments are brought together within the same molecule (31–33).

There is also a family of triplexes built from a single strand and a hairpin. Two types of arrangements are possible for such structures (34, 35): (*a*) a canonical hairpin, formed by two self-complementary DNA segments, is involved in Hoogsteen hydrogen bonding with a single strand (Figure 1*C,H*), and (*b*) a "hairpin" containing a homopurine (for YR*R) or homopyrimidine (for YR*Y) mirror repeat is involved in both Watson-Crick and Hoogsteen basepair formation in a triplex (Figure 1*D,I*). A peculiar modification of this scheme was described in Refs. 36 and 37, where a short circular oligonucleotide could be used for triplex formation instead of a hairpin (Figure 1*E,J*). Such a triplex-forming oligonucleotide is of particular interest for DNA targeting in vivo (see below), since circular oligonucleotides are not substrates for degradation by exonucleases.

The structures in Figure 1 are intentionally ambiguous with regard to the 5′ and 3′ ends of polynucleotide chains. In fact, all these structures may exist as two chemically distinct isoforms differing in relative chain polarity. The comparative stability of the two isoforms is poorly known. Very recent data presented in Refs. 38 and 39 indicate that their free energies may differ by 1.5–2.0 kcal/mol and may depend on the loop sequence. The two isoforms may also differ topologically (see below).

So far, we have considered triplexes with their duplex part consisting of purely homopurine and homopyrimidine strands (the influence of individual mismatched triads is discussed below). It has become clear recently, however, that both sequence requirements and chain polarity rules for triplex formation can be met by DNA target sequences built of clusters of purines and pyrimidines (40–43) (see Figure 1*K–M*). The third strand consists of adjacent homopurine and homopyrimidine blocks forming Hoogsteen hydrogen bonds with purines on alternate strands of the target duplex, and this strand switch

preserves the proper chain polarity. These structures, called alternate-strand triplexes, have been experimentally observed as both intra- (42) (Figure 1*M*) and intermolecular (41, 43) (Figure 1*K,L*) triplexes. These results increase the number of potential targets for triplex formation in natural DNAs somewhat by adding sequences composed of purine and pyrimidine clusters, although arbitrary sequences are still not targetable because strand switching is energetically unfavorable. Preliminary estimates give the minimal length of a cluster in an alternate-strand triplex as between 4 and 8 (44). A peculiar feature of alternate-strand triplexes is that two different sequences of the third strand fulfill the requirements for triplex formation for a single duplex target (Figure 1*K,L*). For a few studied targets, the efficiency of triplex formation by the two variants was quite different. Strand switching in the direction 3′-R_n-Y_n-5′ along the third strand was more favorable than 3′-Y_n-R_n-5′ (44).

Hybrid triplexes consisting of both DNA and RNA chain are less studied and only for YR*Y triplexes. Eight combinations of RNA and DNA chains within a triplex are possible in principle, and the relative stability of each was studied (45–47). The results from different groups differ substantially, for reasons that are yet to be understood. (Though they may be attributed to differences in sequences and/or the ambient conditions.) However, all these studies show consistently that triplexes are more stable when DNA represents the central homopurine strand than when RNA does. Affinity cleavage data also indicate that the orientation of chemically homologous chains in hybrid triplexes is antiparallel.

Fine Structure of DNA Triplexes

The structural features of DNA triplexes have been studied using such diverse techniques as chemical and enzymatic probing, affinity cleavage, and electrophoresis, all of which have provided insight into the overall structure of triplexes: (*a*) The third strand lies in the major groove of the duplex, as is deduced from the guanine N7 methylation protection (48, 49); (*b*) the orientation of the third strand is antiparallel to the chemically homologous strand of the duplex (11, 26); and (*c*) the duplex within the triplex is noticeably unwound relative to the canonical B-DNA (12). However, the fine structure of triplexes could not be elucidated at atomic resolution without more direct structural methods based on X-ray diffraction and NMR.

The first attempt to deduce the atomic structure of a poly(dT)•poly(dA)•poly(dT) triplex using X-ray fiber diffraction was performed in 1974 (50). Two important parameters of the triple helix, an axial rise equal to 3.26 Å and a helical twist of 30°, were directly determined from the fiber diffraction patterns. However, in an attempt to fit experimental data with atomic models, the

Table 1 Computed averages for various helical parameters

	Triplexes					
	YR*Y		YR*R			
	X-ray	NMR	X-ray	NMR	B-DNA	A-DNA
Axial rise (Å)	3.3	3.4	—	3.6	3.4	2.6
Helical twist (°)	31	31	—	30	36	33
Axial displacement (Å)	−2.5	−1.9	—	−1.9	−0.7	−5.3
Glycosidic torsional angle	anti	anti	—	anti	anti	anti
Sugar pucker	C2' endo	C2' endo	—	C2' endo	C2' endo	C3' endo

duplex within the triplex was erroneously concluded to adopt the A conformation (50). Other studies of the atomic structure of triplexes have only recently corrected this widely accepted conclusion. NMR data (51) (see below) and infrared spectroscopy (52) convincingly demonstrated that the sugar pucker in all three strands within the triplex is of the S-type (a characteristic for B-DNA rather than A-DNA). It appeared that a B-like structure could nicely explain the original fiber diffraction data (Table 1). Moreover, this structure is stereochemically more favorable than is the original (53).

Further sophisticated NMR studies have examined inter- and intramolecular triplexes of both YR*Y and YR*R types. This data unambiguously supported the major features of the YR*Y triplexes discussed above: (*a*) a requirement for cytosine protonation (54–56); (*b*) Hoogsteen basepairing of the third strand (32, 57); and (*c*) antiparallel orientation of the two pyrimidine strands (32, 57). The atomic structure of the triple helix, summarized in Table 1, was also determined (51, 58). The values of all major parameters determined by NMR are very close to those determined by fiber diffraction (53). Most significantly, the deoxyribose conformation of all strands in the triplex corresponds to an S-type (C2′-endo) pucker (51). It is clear from Table 1 that the duplex within the triplex adopts a B-like configuration; the helical twist in the triplex, however, is significantly smaller than that for B-DNA.

NMR studies of the YR*R triplexes (33, 59) showed that their overall structure is similar to that of YR*Y triplexes. The important difference, however, is that reverse Hoogsteen basepairing of the third strand (as in Figure 3) was convincingly demonstrated. The helical parameters presented in Table 1 are close to those of YR*Y triplexes and suggest the formation of an unwound B-like structure. A peculiar feature of the YR*R triplexes consisting of CG*G and TA*T triads is the concerted changes in the axial rise and helical twist along the helix axis (60). This was attributed to the lack of isomorphism between CG*G and TA*T triads, discussed above.

H Form

Although the canonical Watson-Crick double helix is the most stable DNA conformation for an arbitrary sequence under usual conditions, some sequences within duplex DNA are capable of adopting structures quite different from the canonical B form under negatively superhelical stress (reviewed in 61). One of these structures, the H form, includes a triplex as its major structural element (Figure 4). Actually, there is an entire family of H-like structures (see 20 for a comprehensive review).

The term "H form" was proposed in a study of a cloned sequence from a spacer between the histone genes of sea urchin (62). It contained a $d(GA)_{16}$ stretch hypersensitive to S1 nuclease. Such S1-hypersensitive sites had been anticipated previously to adopt an unusual structure (63, 64), and numerous hypotheses had been discussed in the literature (63, 65–71). Using 2-D gel electrophoresis of DNA topoisomers (see 20), a structural transition was demonstrated without enzymatic or chemical modification (62). The pH dependency of the transition was remarkable: At acidic pH the transition occurs under low torsional tension, while at neutral pH it is almost undetectable. Because pH dependence had never been observed before for non-B-DNA conformations (cruciforms, Z-DNA, bent DNA, etc), the investigators concluded that a novel DNA conformation was formed. The structure was called the H form because it was clearly stabilized by hydrogen ions, i.e. it was a protonated structure.

The H form model proposed in Ref. 8 (Figure 4*A*) consists of an intramolecular triple helix formed by the pyrimidine strand and half of the purine strand, leaving the other half of the purine strand single stranded. As Figure 4*A* shows, this structure is topologically equivalent to unwound DNA. Two isoforms of H form are possible: one single stranded in the 5′ part of the purine strand and the other single stranded in the 3′ part. The existence of single-stranded purine stretches in H-DNA explains its hyperreactivity to S1 nuclease. Canonical TA*T and CG^*C^+ base triads stabilize the triple helix (Figure 2). The protonation of cytosines is crucial for the formation of CGC^+ base triads, which explains the pH dependency of the structural transition.

The H-DNA model predicts that a homopurine-homopyrimidine sequence must be a mirror repeat to form H-DNA. It was convincingly demonstrated in Ref. 9 that mirror repeats indeed adopt the H form, while even single-base violations of the mirror symmetry significantly destabilize the structure. Chemical probing of H-DNA using conformation-specific DNA probes (reviewed in 72) provided final proof of the H-form model (25, 48, 49, 73–76). Notably, these studies revealed that different sequences preferentially adopt only one of the two possible isomeric forms of H-DNA, the one in which the 5′ part of the purine strand is unstructured.

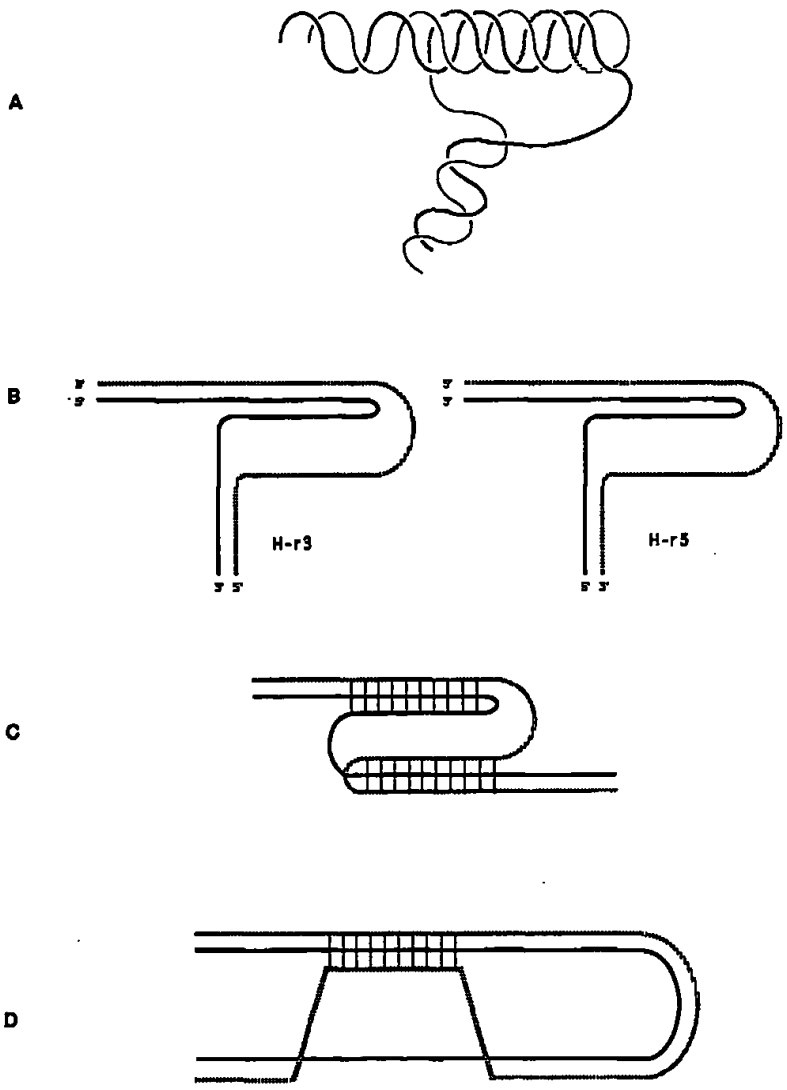

Figure 4 H-DNA menagerie. *A.* H-DNA model. Bold line, homopurine strand; thin line, homopyrimidine strand; dashed line, the half of the homopyrimidine strand donated to the triplex. *B.* Two isoforms of *H-DNA. *C.* Nodule DNA. *D.* Tethered loop. In *B–D*, solid line, homopurine strand; stippled line, homopyrimidine strand.

The structural features responsible for the difference between the two isoforms have been identified in Ref. 77. The isoform with the 3′ half of the pyrimidine strand donated to the triplex (designated H-y3) is preferentially formed at physiological superhelical densities. In this isoform, the 5′ portion of the purine strand is single stranded, and its formation is consistent with the chemical probing results described above. The other isoform (in which the 5′ half of the pyrimidine strand is donated to the triplex—designated H-y5) was only observed at low superhelical density. Topological modeling of H-DNA formation showed that the formation of the H-y3 isoform releases one extra supercoil relative to the H-y5 isoform. This explains why H-y3 is favorable at high superhelical density. Recent studies show that the mechanisms underlying preferential isomerization into the H-y3 conformation are more complex. Apparently, the presence of bivalent cations can make the H-y5 isoform preferable (78). What is more surprising, the loop sequence plays an important role in determining the direction of isomerization (79, 80). Systematic studies of factors contributing to isomerization are yet to be done.

H-DNA Menagerie

As for intermolecular triplexes, a menagerie of H-DNA-like structures exists (reviewed in 20). First, intramolecular YR*R triplex, called *H-DNA, was described in Refs. 25 and 81 (Figure 4*B*). This structure is also topologically equivalent to the unwound DNA and requires DNA supercoiling (82). As in intermolecular YR*R triplexes, A can be replaced with T (83) and, at acidic pH, G can be replaced with A (28) in the third strand of *H-DNA. Thus, the sequences adopting the *H form are not necessarily mirror repeated and not even necessarily homopurine-homopyrimidine (see 20 for comprehensive review).

Two isoforms of *H-DNA are possible, designated H-r3 and H-r5 according to which half of the homopurine strand is donated to the triplex (Figure 4*B*). Chemical probing with single-stranded, DNA-specific agents showed that H-r3 isoform is dominant.

As for all YR*R triplexes, the mechanisms of *H-DNA dependence on bivalent cations are unclear. Cation requirements are different for different sequences (20, 25, 27, 81, 84–87). For example, while *H-DNA formed by $d(G)_n \cdot d(C)_n$ sequences is stabilized by Ca^{2+}, Mg^{2+}, and Mn^{2+}, the same structure formed by $d(GA)_n \cdot d(TC)_n$ is formed in the presence of Zn^{2+}, Mn^{2+}, Cd^{2+}, and Co^{2+}. The differences in cation requirements are due to variations in neighboring triads or changes in the GC content or both. Even moderate changes in GC content (from 75% to 63%) switched cation requirement from Mg^{2+} to Zn^{2+} for a particular sequence (22). A Mg^{2+}-to-Zn^{2+} switch was reported to affect the equilibrium between H-r5 and H-r3 isoforms (86) or to substantially modify the *H-structure (87).

A hybrid of H and *H forms was described, called nodule DNA (88, 89) (Figure 4*C*). Nodule DNA is an analog of the intermolecular alternate-strand triplexes described above.

A peculiar H-like structure formed by two distant homopurine-homopyrimidine tracts was described in Ref. 90. It is in a way similar to an early model for S1 hypersensitivity in the human thyroglobulin gene (69). It was found that linear DNA containing both tracts at pH 4.0 and in the presence of spermidine migrates very slowly in an agarose gel. This abnormal electrophoretic mobility was attributed to the formation of a so-called tethered loop (Figure 4*D*). In this structure, the homopyrimidine strand of one stretch forms a triplex with a distant stretch, while its complementary homopurine strand remains single stranded. Supporting this model, it was found that the addition of excess homologous homopyrimidine, but not homopurine, single-stranded DNA prevented loop formation. Though the mechanism of tethered loop formation is not self-evident, it is allowed topologically. Chemical probing is required to prove the existence of this structure definitively.

Specificity of Triplex Formation

The specificity and stringency of triplex formation (35) has attracted serious attention for two reasons. First, the formation of triplexes is limited to the homopurine-homopyrimidine sequences or to sequences composed of adjacent oligopurine/oligopyrimidine clusters. This major limitation to the biological and theurapeutic applications of triple-helical DNAs prompted an extensive search for DNA bases that could be incorporated into the third strand of a triplex in order to recognize thymines or cytosines in the otherwise homopurine strand of the duplex. Secondly, accurate knowledge of the specificity of third-strand recognition for perfect homopurine-homopyrimidine sequences is necessary in order to target natural DNAs.

The quest for such knowledge stimulated the study of non-orthodox triads. So far most of the data have been collected for YR*Y triplexes, including all 14 noncanonical triads (other than CG*C and TA*T). One approach was to analyze the influence of mismatched triads on H-DNA formation using 2-D gel electrophoresis (91). Stability of mismatched triads in intermolecular triplexes was studied using affinity cleavage (92), melting experiments (93, 94), and NMR (95). These studies agreed that although single mismatches could be somewhat tolerated, each mismatch significantly disfavored triplex. The mismatch energies were within the range of 3–6 kcal/mol, i.e. similar to the cost of B-DNA mismatches. Thus, homopyrimidine oligonucleotides form triplexes with target sequences at a specificity comparable to that seen in Watson-Crick complementary recognition.

High sequence specificity of third-strand recognition of homopurine-homopyrimidine sequences in the duplex makes TFOs very attractive candidates for

targeting genomic DNA. Supporting this conclusion, homopyrimidine TFOs equipped with Fe•EDTA have been demonstrated to cleave unique sites in yeast (96, 97) and human (98) chromosomes. They were also found to be convenient tools for affinity capture of human genomic targets (99).

However, studies widely disagreed on the relative stability of individual noncanonical triads. For example, the AT*G triplet was shown to be the most favorable in studies of intermolecular triplexes (92, 93), but it is not among the best for H-DNA (91). This contradiction could be due to the different triplex-forming sequences studied by different groups, since heterogeneity in stacking interactions within a triple helix must seriously affect its stability. This idea was recently supported by NMR studies of the AT*G triad (58, 100). It was found that guanine in this triplet is tilted out of the plane of its target AT basepair to avoid a steric clash with the thymine methyl group. This causes a favorable stacking interaction between this guanine and the thymine flanking it from the 5′-side, which is likely to be a major determinant of AT*G triplet stability. This also explains the differences between the inter- and intramolecular triplex studies: In the first case, guanine was flanked by thymine on the 5′ end (92), while in the second case, it was flanked by a cytosine (91). Thus, the favorable stacking interaction was absent in the intramolecular triplex, and the AT*G triad was relatively unstable. Recently, it was shown directly that replacement of the TA*T triad on the 5′ side of guanine with a CG*C triad reduces the stability of TA*G triplet (101). The clear message from these results is that the influence of nearest neighbors on triad stability must be studied to better understand the duplex-to-triplex transition. This doughty goal is not yet achieved.

Notwithstanding the difficulties discussed above, empirical rules for targeting imperfect homopurine-homopyrimidine sequences were suggested in Ref. 102. If the homopurine strand of a duplex is interrupted by a thymine or cytosine, it must be matched by a guanine or thymine, respectively, in the third strand. However, this expansion of the third-strand recognition code is premature, as was recently addressed (103). The GC*T triad, though reasonably stable, is dramatically weaker than the canonical TA*T triad. Thus, a TFO containing a thymine, intended to interact with a cytosine in the target, would bind significantly better to a different target containing adenine in the corresponding position. In the AT*G case, the triad specificity is high, but the affinity of the G for the TA pair is only modest.

Another approach to overcoming the homopurine-homopyrimidine target requirements is to incorporate artificial DNA bases within the third strand. Several studies found that non-natural bases, such as 2′-deoxynebularine or 2′-deoxyformycin A and others, may form very stable triads with cytosines and thymines intervening the homopurine strand (94, 104, 105). It is yet to be seen if the specificity and stringency of such complexes is sufficient.

Limited data are available on the mismatched triads in YR*R triplexes. By use of affinity cleavage experiments, all 13 noncanonical triplets (all combinations except CG*G, TA*A, and TA*T) were shown to disfavor triplex formation (106). The only notable exception is the CG*A triad, which is favorable under acidic pH due to the protonation of its adenine (28). Much as with homopyrimidine TFOs, purine-rich TFOs can be used specifically to target homopurine-homopyrimidine sequences in natural DNAs.

Stabilization of Triplexes

The stabilization of DNA triplexes is particularly important for any possible biological applications. As discussed above, the YR*Y triplexes are formed under acidic pH, while YR*R triplexes require millimolar concentrations of bivalent cations. Physiological pH, however, is neutral, and a high concentration of unbound bivalent cations in a cell is unlikely. Thus, numerous studies have been aimed at the stabilization of DNA triplexes at physiological conditions.

Most of the YR*Y triplexes studies have been concentrated on overcoming pH dependency. The most promising results show that polyamines, specifically spermine and spermidine, favor both inter- and intramolecular YR*Y triplexes under physiological pH (11, 107, 108). The stabilizing effect is likely due to decreased repulsion between the phosphate backbones after binding to polyamines, overcoming the relatively high density of a negative charge in triplexes. The millimolar polyamine concentrations found in the nuclei of eukaryotic cells (reviewed in 109) raise the hope for triplexes in vivo.

The requirement for cytosine protonation could be overcome by several chemical means. The incorporation of 5-methylcytosines instead of cytosines in TFOs increases the stability of YR*Y triplexes at physiological pH (110, 111), but more detailed study found that this effect is relatively small (the apparent methylation-induced ΔpK_a is only 0.5) (112). Another solution is to substitute cytosines in the third strand with non-natural bases that do not require protonation for Hoogsteen hydrogen bond formation. Indeed the substitution of cytosines with N^6-methyl-8-oxo-2-deoxyadenosines (113), pseudoisocytidines (114), 7,8 dihydro-8-oxoadenines (115), or 3-methyl-5-amino-1H pyrazolo [4.3-d] pyrimidin-7-ones (116) led to pH-independent triplex formation.

Intermolecular triplexes could be additionally stabilized if the third strand represented an oligodeoxynucleotide-intercalator conjugate. This was first demonstrated for a homopyrimidine oligonucleotide linked with an acridine derivative (10) and later shown for other oligonucleotides and intercalators (117–119). The stabilization is due to the intercalation of a ligand into DNA at the duplex-triplex junction. For reasons that are yet unclear, the most stable complex is formed when the intercalator is attached to the 5′ end of the TFO.

Particularly promising for gene targeting is an oligonucleotide-psoralen conjugate, as near-UV irradiation of a triplex formed by such a conjugate leads to crosslink formation, making the triplex irreversible (120).

An independent line of research has sought for triplex-specific ligands. One such ligand, a derivative of benzo[e]pyridoindole (BePI), has been described in Refs. 121 and 122. BePI shows preferential intercalation into a triple- rather than double-helical DNA, thus greatly stabilizing triplexes (122). Another promising triplex-binding ligand is coralyne (123).

It should be emphasized that, to be prospective drugs for gene targeting, TFOs must meet two requirements: They must bind their targets relatively strongly and not target other sequences. If a TFO has very strong affinity to its target, it can also bind to a site with one or even more mismatches. This should be especially true for non–sequence-specific stabilization of triplexes with intercalating drugs attached to TFOs. Therefore, increased stability inevitably entails decreased selectivity of the TFO. It is not at all accidental that the spectacular demonstration of sequence-selective cutting of genomic DNA with TFOs was achieved under conditions of extremely weak binding of the TFO to its target site (96–98). Systematic experimental study of sequence selectivity of all modified TFOs mentioned above is still lacking. However, it is obvious that these modified TFOs should exhibit poorer selectivity than do the original TFOs.

The stabilization of intramolecular triplexes could be achieved in several ways, the most obvious of which is to increase the negative superhelical density, since the formation of H-DNA releases torsional stress. As is discussed below, the increase of negative supercoiling does provoke triplex formation in vivo. The polyamine stabilization of H-DNA at physiological pH has already been mentioned.

A less obvious way of stabilizing H-DNA, called kinetic trapping, was described (124). It was found that oligonucleotides complementary to the single-stranded homopurine stretch in H-DNA stabilized H-DNA under neutral pH, where H-DNA alone rapidly reverts to the B conformation.

Peptide Nucleic Acid (PNA)

PNA is the prototype of an entire new class of TFO-based drugs that interact with DNA in a manner unlike that of ordinary TFOs. PNA (Figure 5*A*) was designed in the hope that such an oligonucleotide analog containing normal DNA bases with a polyamide (i.e. proteinlike) uncharged backbone would form triplexes with double-stranded DNA (dsDNA) much more efficiently than do the regular TFOs (14).

Instead of forming triplexes with duplex DNA, the first studied homothymine PNA oligomer, PNA-T_{10}, opened the DNA duplex in A_n/T_n tracts, forming an exceptionally strong complex with the A strand and displacing the T strand

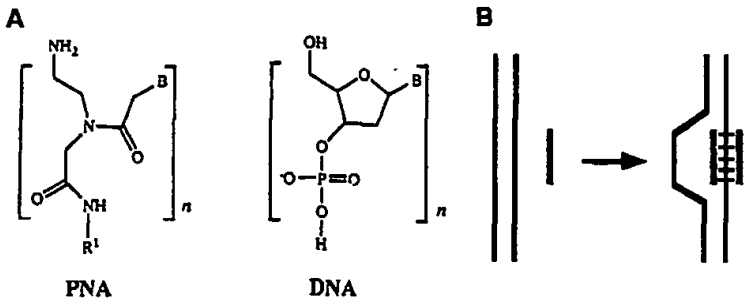

Figure 5 *A.* The chemical structures of PNA and DNA. *B.* P-loop formation. Bold line, DNA; stippled line, PNA.

(15, 125, 126). At the same time, model experiments with complexes formed between PNA oligomers and oligonucleotides revealed that, while PNA/DNA heteroduplexes are not much more stable under ordinary conditions than are DNA/DNA homoduplexes (127), two homopurine PNA oligomer molecules form exceptionally stable triplexes with the complementary homopurine oligonucleotide (128, 129).

These results strongly suggest an unusual mode of binding between the synthetic analog and dsDNA. Namely, two homopyrimidine PNA molecules displace the duplex DNA pyrimidine strand and form a triplex with the purine strand of DNA (15, 16, 130, 131). These complexes are called the P-loops (Figure 5*B*).

The P-loop is a radically different complex than that formed between duplex DNA and ordinary TFOs. Although the fact of $(PNA)_2$/DNA triplex formation during the strand-displacement reaction has been convincingly proven (16, 130, 131), the mechanism of P-loop formation remains to be elucidated. The available data indicate that the reaction most probably proceeds via a short-lived intermediate, which consists of one PNA molecule complexing with the complementary DNA strand via Watson-Crick pairing. This intermediate is formed due to thermal fluctuations (breathing) of the DNA duplex (132, 133). It is very unstable and would dissociate if it were not fixed by the second PNA oligomer in a $(PNA)_2$/DNA triplex leading to P-loop formation (see Figure 5*B*). This triplex is remarkably stable.

PNA forms much more stable complexes with dsDNA than do regular oligonucleotides. This makes PNA very promising as an agent for sequence-specific cutting of duplex DNA (16), for use in electron-microscopy mapping of dsDNA (15), and as a potential antigene drug (134, 135), as PNA is

remarkably stable in biological fluids in which normal peptides and oligonucleotides are quickly degraded (136).

However, serious limitations for various applications of PNA still remain. P-loop formation proceeds through a significant kinetic barrier and strongly depends on ionic conditions (15, 16, 125, 126). This dependency, if not bypassed, poses significant limitations on possible sequence-specific targeting of dsDNA by PNA under physiological conditions. Although the stringency of $(PNA)_2$/DNA triplexes is not yet known, PNA should still target predominately homopurine-homopyrimidine regions, just as do regular TFOs.

BIOCHEMISTRY OF TRIPLEXES

Formation and Possible Functions of H-DNA In Vivo

As is true for other unusual DNA structures, such as cruciforms, Z-DNA, and quadruplexes, the biological role of H-DNA is yet to be established. Two important problems must be addressed: (*a*) Can H-DNA be formed in cells in principle? (*b*) In which biological process if any is H-DNA involved? Recently it became clear that the answer to the first question is yes. There are currently many hypotheses on the role of H-DNA in DNA replication, transcription, and recombination, but more studies are needed to answer the second question.

Sequences that can form H-DNA are widespread throughout the eukaryotic genomes (137, 138) but are uncommon among eubacteria. However, direct detection of H-DNA in eukaryotic cells is very difficult because of the complexity of genomic DNA. Therefore, most of the studies on the detection of H-DNA in vivo exploited *Escherichia coli* cells bearing recombinant plasmids with triplex-forming inserts as convenient model systems. Chemical probing of intracellular DNA proved helpful for the detection of H-DNA in vivo. Certain chemicals, such as osmium tetroxide, chloroacetaldehyde, and psoralen, give a characteristic pattern of H- or *H-DNA modification in vitro. Conveniently, they can also penetrate living cells. Thus, the general strategy for detecting H-DNA in vivo was to treat *E. coli* cells with those chemicals, isolate plasmid DNA, and locate modified DNA bases at a sequence level. The coincidence of modification patterns in vitro and in vivo basically proved the formation of the unusual structure in the cell.

Using this approach, the formation of both H- and *H-DNA was directly shown (139–141). The corresponding studies were reviewed in Ref. 20, but we briefly summarize the major findings. All these studies agreed that the level of DNA supercoiling in vivo is the major limiting factor in the formation of these structures. Though transient formation of H-DNA was observed in normal exponentially growing *E. coli* cells (141), formation of H-DNA was much more pronounced when intracellular DNA supercoiling increased, due to mu-

tations in the gene for Topo I (141) or due to treatment of cells with chloramphenicol (139, 140). Environmental conditions during *E. coli* growth also significantly contributed to the appearance of triplexes. H-DNA formation was greatly enhanced when cells were growing in mildly acidic media, which somewhat decreased intracellular pH (139, 141) while *H-DNA was observed in cells growing in media with a high concentration of Mg^{2+} ions (140). Neither result is surprising, because H-DNA is stabilized by protonation while *H-DNA is stabilized by bivalent cations.

Besides the steady-state level of DNA supercoiling, determined by the balance of DNA gyrase and Topo I (reviewed in 142), the local level of supercoiling strongly depends on transcription. During the process of polymerization the RNA polymerase creates domains of high negative and positive supercoiling upstream and downstream of it, respectively (143), which may influence the formation of unusual DNA structures (144, 145). Chemical probing of intracellular DNA demonstrated transcriptionally driven formation of *H-DNA within long $d(G)_n \cdot d(C)_n$ stretches located upstream of a regulated promoter in an *E. coli* plasmid (146). Remarkably, the formation of *H-DNA stimulated homologous recombination between direct repeats flanking the structure. Thus, this work shows the formation of *H-DNA under completely physiological conditions in a cell, and implicates it in the process of recombination.

The only data on triplex DNA detection in eukaryotic cells were obtained using antibodies against triple-helical DNA (147). These antibodies were found to interact with eukaryotic chromosomes (148, 149).

Many ideas have been proposed involving H-DNA in such basic genetic processes as replication and transcription. The hypothesis regarding H-DNA in replication is based on the observation that triplex structures prevent DNA synthesis in vitro. On supercoiled templates containing *H-DNA, DNA synthesis prematurely terminates. The location of the termination site is different for different isoforms of *H-DNA, but it always coincides with the triplex boundaries as defined by chemical probing (83).

More peculiarly, H-like structures can be formed in the process of DNA polymerization and efficiently block it. Two such mechanisms were demonstrated experimentally (Figure 6*A*,*B*). It was found that $d(GA)_n$ or $d(C\text{-}T)_n$ inserts within single-stranded DNA templates cause partial termination of DNA polymerases at the center of the insert (21, 150). It was suggested that when the newly synthesized DNA chain reaches the center of the homopolymer sequence, the remaining homopolymer stretch folds back, forming a stable triplex (Figure 6*A*). As a result, the DNA polymerase finds itself in a trap and is unable to continue elongation.

In open circular DNA templates, H-like structures are absent due to the lack of DNA supercoiling. It was shown, however, that T7 DNA polymerase ter-

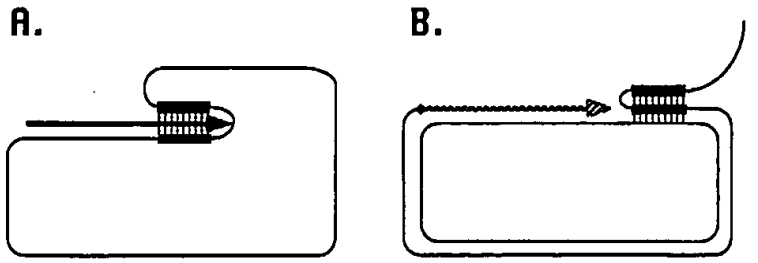

Figure 6 DNA polymerase–driven triplex formation blocks polymerization. Black boxes, the two halves of a homopurine-homopyrimidine mirror repeat involved in the formation of an intramolecular triplex; striated arrow, the newly synthesized DNA chain. *A*. Single-stranded DNA template. *B*. Double-stranded DNA template.

minated exactly at the center of *H-forming sequences. This was observed when the pyrimidine-rich but not the purine-rich strand served as a template (22). To explain this one must remember that DNA synthesis on double-stranded templates is possible due to the ability of many DNA polymerases to displace the nontemplate DNA strand (reviewed in 151). The displaced strand may fold back, promoting the formation of an intramolecular triplex downstream of the replication fork at an appropriate sequence. Conditions for DNA synthesis in vitro—i.e. neutral pH and high magnesium concentration—are optimal for the formation of YR*R triplexes. Thus, the displacement of the purine-rich (but not the pyrimidine-rich) strand provokes triplex formation which, in turn, leads to termination of DNA synthesis (Figure 6*B*).

There are only fragmentary data on the role of H motifs in the regulation of replication in vivo. Several homopurine-homopyrimidine inserts were shown to decrease the efficiency of Simian virus 40 (SV40) DNA replication (152, 153). Quite recently, the pausing of the replication fork in vivo within a $d(GA)_n{\cdot}d(TC)_n$ insert in SV40 DNA was demonstrated directly using a technique called two-dimensional neutral/neutral gel electrophoresis (23). Though these data make the idea of H-DNA involvement in the regulation of replication promising, it is far from proven. Future studies are crucial for the evaluation of this hypothesis.

Numerous studies concerned the possible role of H-DNA in transcription. Deletion analysis of various promoters—including *Drosophila hsp*26 (154, 155); mouse c-Ki-*ras* (156) and TGF-β3 (157); human EGFR (158), *ets*-2 (159), IR (160), and c-*myc* (161, 162); and others—showed that homopurine-homopyrimidine stretches are essential for promoter functioning.

These sequences serve as targets for nuclear proteins, presumably transcrip-

tional activators. Several homopurine-homopyrimidine DNA-binding proteins were described, including BPG1 (163), NSEP-1 (164), MAZ (165), nm23-H2 (166), PYBP (167), Pur-1 (168), etc. Peculiarly, these proteins often bind preferentially to just one strand of the H motifs. For example, a number of mammalian proteins specifically recognize homopurine-homopyrimidine sequences in the double-helical state as well as the corresponding homopyrimidine single strands (164, 167, 169, 170). This unusual binding pattern may dramatically influence the equilibrium between different DNA conformations in the promoter in vivo.

However, the importance of the H structure for transcription was questioned in several studies. One approach is to analyze the influence of point mutations within H motifs that destroy or restore H-forming potential on the promoter's activity. No such correlation was observed for *Drosophila hsp26* (155) and mouse c-Ki-*ras* (171) promoters. The situation with the c-*myc* promoter is more complex, since it is unclear if the canonical H-DNA or some other structure is formed even in vitro (172). Mutational analysis of the promoter gave contradictory results, with one group claiming the existence (173) and another the lack (174) of a correlation between structural potential and promoter strength. Another approach to detecting H-DNA in eukaryotic promoters is direct chemical probing followed by genomic sequencing. So far, this has only been done for the *Drosophila hsp26* gene, and H-DNA was not observed (155).

It is hard to completely rule out the role of H-DNA in transcription based on the above results. First, it is quite possible that the structural peculiarities of promoter DNA segments may affect the interaction between promoter DNA and specific regulator proteins. The features of homopurine-homopyrimidine DNA-binding proteins described above as well as a report about the partial purification of a triplex-binding protein (175) indirectly support this idea. A study in which the influence of $d(G)_n$ stretches of varying length on the activity of a downstream minimal promoter was analyzed additionally supports this hypothesis (176). A clear reverse correlation between the ability of a stretch to form the *H configuration in vitro and its ability to activate transcription in vivo was observed. It was concluded, therefore, that short $d(G)_n$ stretches serve as binding sites for a transcriptional activator, while longer stretches adopt a triplex configuration, which prevents activator binding. Secondly, negative data on the role of H-DNA in transcription were obtained in transient assays, while it can actually work at a chromosome level. Indeed, H motif in the *Drosophila hsp26* gene was found to affect the chromatin structure (177, 178).

Despite the wealth of data and hypotheses, there is no direct evidence that the structural features of H motifs are involved in transcriptional regulation in vivo, and further studies are required to address this issue.

Targeting Basic Genetic Processes Using TFOs

Highly sequence-specific recognition of double-helical DNAs by TFOs is the basis of an antigene strategy (reviewed in 13). The idea is that binding of a TFO to a target gene could prevent its normal functioning. Most studies of this strategy concerned the inhibition of transcription; the studies were inspired in part by the existence of functionally important homopurine-homopyrimidine stretches in many eukaryotic promoters (see the previous section), which are appropriate targets for TFOs. The antigene strategy could potentially lead to rational drug design. Very convincing data on the inhibitory effects of TFOs were obtained in various in vitro systems. There are also preliminary indications that TFOs may function in vivo as well.

The first stage that is affected by TFOs is the formation of an active promoter complex. Pioneering results were obtained for the human c-*myc* promoter, where it was found that the binding of a purine-rich TFO to the imperfect homopurine-homopyrimidine sequence 125 basepairs (bp) upstream of the P1 promoter start site blocks its transcription in vitro (179). The TFO's target is important for c-*myc* transcription, serving as a binding site for a protein(s), presumably a transcriptional activator (161, 162). At least two candidate genes coding for proteins that bind to this target have been cloned and sequenced (164, 166). Similar observations were made for the methallothionein gene promoter. In this case a homopyrimidine oligonucleotide formed a triplex with the upstream portion of the promoter, preventing the binding of the transcriptional activator Sp1 (111). This in turn drastically reduced the promoter's activity in a cell-free transcription system (179a). TFOs were also shown to prevent SP1 binding to the human DHFR (180) and *H-ras* (181) promoters. Finally, a triplex-forming oligonucleotide-intercalator conjugate was shown to act as a transcriptional repressor of the interleukin-2 receptor α gene in vitro (182), preventing the binding of the transcriptional activator NFκB. In all these cases TFOs efficiently blocked the access of the transcription factors to their binding sites.

TFOs also inhibit initiation of transcription by RNA polymerases. The pBR322 *bla*-gene contains a 13-bp homopurine-homopyrimidine target just downstream of the transcriptional start site. A 13-mer homopyrimidine oligonucleotide forming an intermolecular triplex with this target hindered initiation of transcription by *E. coli* RNA polymerase in vitro (183). Independent studies showed that this is also the case for T7 RNA polymerase (184).

Finally, eukaryotic RNA polymerase II transcription was followed in vitro from the adenovirus major late promoter (185). The transcribed portion of DNA contained a 15-bp homopurine-homopyrimidine tract that formed an intermolecular triplex with the homopyrimidine TFO. When added prior to RNA polymerase, the TFO truncated a significant portion of the transcripts.

Thus, TFOs can block transcription at different stages: promoter complex formation, initiation, and elongation. This appears to be true for both pro- and eukaryotic RNA polymerases. TFOs can be considered to be artificial repressors of transcription (186).

There is a growing number of indications that TFOs may act as repressors of transcription in cell cultures as well. The most convincing results so far were obtained for the interleukin-2 receptor α promoter (182, 187). Homopyrimidine TFOs were designed to overlap a target site and prevent binding of the transcriptional activator NFκB. They were conjugated with acridine to stabilize the triplexes, or psoralen to make triplex formation irreversible after UV irradiation. The plasmid bearing the reporter gene under the control of the IL-2Rα promoter was cotransfected with these TFOs in tissue cultures, where it was shown that TFOs block promoter activity in vivo. Particularly strong inhibition was observed after UV irradiation of cells transfected with psoralen conjugates. In the latter case, chemical probing directly demonstrated the formation of intermolecular triplex in vivo. A similar cotransfection approach was also used to target Interferon Responsive Elements in vivo (188).

A different approach was used in several studies where purine-rich TFOs were added to the growth media of cells containing target genes. To prevent oligonucleotides from degrading, their 3′ ends were protected by an amino group (189). Such oligonucleotides accumulated within cells and could be recovered in intact form. Partial transcriptional inhibition of human c-*myc* and IL2Rα genes by such TFOs has been reported (189, 190). Similar effects were observed for human immunodeficiency virus (HIV) transcriptional inhibition in chronically infected cell lines (191). Using cholesterol-substituted TFOs, the progesterone-responsive gene has also been inhibited (192). Though the inhibitory effect was never more than 50%, it is quite remarkable considering that a short oligonucleotide must find its target in an entire genome and prevent its proper interaction with cellular transcriptional machinery. Note, however, that in none of those cases was the formation of triplexes directly demonstrated. Other mechanisms of oligonucleotide-caused transcriptional inhibition must be ruled out in the future.

The use of TFOs for DNA replication inhibition is less studied. In vitro formation of putative intramolecular triplexes or H-like triplexes (see Figure 1) on single-stranded DNA templates traps many different DNA polymerases (22, 193). Purine-rich TFOs are particularly efficient even against such processive enzymes as T7 DNA polymerase and thermophilic Taq and Vent polymerases, because the conditions of DNA synthesis in vitro are favorable for YR*R triplexes. Pyrimidine-rich TFOs must be additionally crosslinked to the target to cause inhibition (194). TFOs also block DNA polymerases on double-stranded templates (195). The inhibition of DNA synthesis in vitro was observed not only when triplexes blocked the path of DNA polymerase, but

also when a polymerization primer was involved in triplex formation (193). Single-stranded DNA-binding protein (SSB protein) helped DNA polymerases partially overcome the triplex barrier, but with an efficiency dramatically dependent on the triplex configuration.

Though these observations make TFOs promising candidates for trapping DNA replication in vivo, there are almost no experimental data regarding this. The only published data concern the use of an octathymidilate-acridine conjugate, which binds to a $d(A)_8$ stretch in SV40 DNA adjacent to the T antigen–binding site. In vivo it partially inhibits SV40 DNA replication, presumably by interfering with the DNA binding or with unwinding activities of the T antigen (196).

The major problem with the use of TFOs is in matching high sequence selectivity with binding that is sufficiently strong to interfere with genetic processes. Under physiological conditions, TFOs bind weakly to their targets, which by itself favors a high sequence selectivity. However, to significantly affect genetic processes, the TFO must be rather long, which limits the number of potential targets, as such long homopurine-homopyrimidine stretches are infrequent.

Three-Stranded DNA Complexes in Homologous Recombination

In this section we briefly discuss a still poorly understood three-stranded DNA complex, formed by RecA protein and, possibly, recombinant proteins from other sources. RecA protein is well known to exhibit many enzymatic activities essential for recombination (reviewed in 18, 197). The main function of RecA protein in recombination is to exchange single-stranded DNA (ssDNA) strand with its homolog in dsDNA. The sequential stages of this reaction are: (*a*) cooperative assembly of RecA protein molecules on the ssDNA, leading to the formation of a right-helical nucleoprotein filament called the presynaptic complex, (*b*) synapsis, i.e. the formation of a complex between this filament and the homologous dsDNA, and (*c*) the actual strand exchange, which requires ATP hydrolysis. Strand exchange proceeds in only one direction: The displacement of a linear single-stranded product starts from its 5′ end.

The synapsis step requires searching for homology between the presynaptic filament and the target dsDNA. One way to do so is to use Watson-Crick complementarity rules. However, this requires a partial strand separation of the dsDNA, resulting in the formation of a so-called D-loop. In this structure, one of the DNA strands of the duplex is displaced, while the other is involved in Watson-Crick pairing with incoming ssDNA. An alternative, very attractive possibility, first postulated in Refs. 17 and 198, does not require dsDNA strand separation and invokes triplex formation. This hypothetical type of DNA triplex was later called "recombination," "parallel," or R-DNA (*19*,

199, 200). These names emphasize two fundamental differences between this hypothetical triplex and the well-characterized orthodox DNA triplexes described in other sections of this chapter. First, chemically homologous DNA strands are parallel in R-DNA but antiparallel in standard triplexes. Secondly, any sequence can adopt an R-DNA conformation, while homopurine-homopyrimidine stretches are strongly preferable in adopting standard triplex structure.

In important experiments on strand exchange between the partially homologous substrates (201, 202), three types of joint molecules were observed. In the case of proximal joints, the area of homology is situated at the 5′ end of the outgoing duplex strand, i.e. both synapsis and strand exchange are possible. For a distal joint (with homology at the 3′ end of the outgoing strand), RecA cannot drive strand exchange. Medial joints contain heterologous regions at both ends of the dsDNA, making strand exchange from any DNA end impossible. Since synaptic complexes were detected in all three cases, it became clear that synapsis and strand exchange are not necessarily coupled. When synaptic complexes—in particular the medial complexes—were treated with DNA crosslinking agents, crosslinks were observed between all three DNA strands involved in the complex (203), indicating a close physical proximity of the three strands.

Analysis of distal joints with very short (38–56-bp) regions of homology showed that they are remarkably stable upon the removal of RecA protein (199). In fact, joint molecules dissociated at temperatures indistinguishable from the melting temperatures of DNA duplexes of the same length and sequence. In spite of its stability, however, the complex did not form spontaneously without recombination proteins. The conclusion was that RecA and related proteins promote the formation of a novel "recombinant" DNA triplex, which otherwise cannot form, presumably due to a kinetic barrier of unknown nature. Independent studies confirmed the extreme stability of deproteinized distal joints with longer regions of homology (204). The basepairing scheme for R-DNA involving triplets for arbitrary DNA sequences was suggested in Refs. 19 and 205. The unique feature of these triplets is the interaction of the third strand with both bases of the Watson-Crick pair.

Although the above data seem to be most consistent with the idea of a "recombination" triplex formation, a careful analysis of three-stranded complexes formed under RecA protein (206) using chemical probing indicates that basepairing in the parental duplex is disrupted. The incoming ssDNA appears to form W-C pairs with the complementary strand of the duplex. It was concluded that the synapsis is accompanied by local unwinding, leading to the formation of D-loop-like structures, rather than the "recombination" triplexes (206). This conclusion was supported by the data that the N7 position of guanines, which is involved in Hoogsteen hydrogen bonding in all known

triplexes in vitro (see Figure 2), is not required for the formation of three-stranded complexes by RecA protein (207).

Thus, the putative triplex between the incoming single strand and the duplex systematically avoids detection. Nevertheless, a more general question remains whether the "recombination" triplexes can be formed in principle, even if they do not play any role in recombination. This kind of triplex has recently been claimed for a postsynaptic complex formed between the outgoing single strand and the duplex yielded as a result of the strand exchange (208).

Quite recently it was suggested that a specifically designed oligonucleotide could fold back to form an intramolecular R-like structure without the assistance of any proteins (209). The main argument is that the thermal denaturation curves are biphasic, which was interpreted as subsequent triplex-to-duplex and duplex-to-single strand transitions. This is hardly a sufficient argument, and data on the chemical and enzymatic probing of such complexes provided in the same study do not support the claim.

In the absence of conclusive evidence, the existence of "recombination" triplexes, or R-DNA, remains doubtful. One of the most uncomfortable questions is the extreme thermal stability of deproteinized distal joints described in Refs. 199 and 204. None of the proposed models can satisfactorily explain this feature. It is totally unclear what is the nature of the kinetic barrier that prevents the formation of R-DNA by dsDNA and homologous oligonucleotide without any protein. It is also unclear why the medial joints, unlike the distal joints, are unstable upon deproteinization (203, 210). Additional concern is possible exonuclease contamination of the RecA protein and SSB protein preparations used for strand transfer reaction. At least in one case, such contamination was admitted to be responsible for the formation of distal junctions (211). In both original papers (199, 204), the authors claimed the lack of nuclease contamination. As shown in (211), however, exonuclease I (ExoI) is enormously activated by SSB protein. As a result, the levels of Exo I required to generate the reverse strand exchange are extremely low (1 molecule of Exo I per 20,000 molecules of RecA protein). In the light of these new findings, it seems possible that distal joints, which were as stable as duplex DNA, might actually be duplexes formed after SSB-activated trace contamination of Exo I digested the nonhomologous strand from its 3′ end.

Even in the absence of a clear understanding of the structure of three-stranded joints promoted by RecA protein, they have already found interesting applications in gene targeting. The first example is called RARE, for RecA-Assisted Restriction Endonuclease cleavage (210). The rationale for this approach is that since RecA protein can form three-stranded complexes between dsDNA and oligonucleotides as short as 15 nucleotides (212), such complexes can be used to block specific methylation sites in dsDNA. After the removal of proteins and consequent dissociation of the three-stranded complexes, cleav-

age by methylase-sensitive restriction endonuclease is limited to the targeted site. Thus, one can cleave large DNAs at a unique site or, using pairs of oligonucleotides, separate specific DNA fragments from the genome.

ACKNOWLEDGEMENTS

We thank our colleagues for sending us valuable reprints and preprints. N Cozzarelli, J Feigon, and B Johnston for comments on the manuscript and R Cox for editorial help. Supported by grant MCB-9405794 from the National Science Foundation to S.M.M.

Literature Cited

1. Felsenfeld G, Davies DR, Rich A. 1957. *J. Am. Chem. Soc.* 79:2023–24
2. Riley M, Maling B, Chamberlin MJ. 1966. *J. Mol. Biol.* 20:359–89
3. Morgan AR, Wells RD. 1968. *J. Mol. Biol.* 37:63–80
4. Lee JS, Johnson DA, Morgan AR. 1979. *Nucleic Acids Res.* 6:3073–91
5. Lipsett MN. 1964. *J. Biol. Chem.* 239: 1256–60
6. Broitman SL, Im DD, Fresco JY. 1987. *Proc. Natl. Acad. Sci. USA* 84:5120–24
7. Cantor CR, Schimmel PR. 1980. *Biophysical Chemistry.* San Francisco: Freeman
8. Lyamichev VI, Mirkin SM, Frank-Kamenetskii MD. 1986. *J. Biomol. Struct. Dyn.* 3:667–69
9. Mirkin SM, Lyamichev VI, Drushlyak KN, Dobrynin VN, Filippov SA, Frank-Kamenetskii MD. 1987. *Nature* 330: 495–97
10. Le Doan T, Perrouault L, Praseuth D, Habhoub N, Decout JL, et al. 1987. *Nucleic Acids Res.* 15:7749–60
11. Moser HE, Dervan PB. 1987. *Science* 238:645–50
12. Lyamichev VI, Mirkin SM, Frank-Kamenetskii MD, Cantor CR. 1988. *Nucleic Acids Res.* 16:2165–78
13. Helene C. 1991. *Anticancer Drug Des.* 6:569–84
14. Nielsen PE, Egholm M, Berg RH, Buchardt O. 1993. In *Antisense Research and Applications*, ed. ST Crooke, B Lebleu, pp. 363–73. Boca Raton, FL: CRC Press
15. Cherny DY, Belotserkovskii BP, Frank-Kamenetskii MD, Egholm M, Buchardt O, et al. 1993. *Proc. Natl. Acad. Sci. USA* 90:1667–70
16. Demidov V, Frank-Kamenetskii MD, Egholm M, Buchardt O, Nielsen PE. 1993. *Nucleic Acids Res.* 21:2103–7
17. Howard-Flanders P, West SC, Stasiak A. 1984. *Nature* 309:215–20
18. West SC. 1992. *Annu. Rev. Biochem.* 61:603–40
19. Zhurkin VB, Raghunathan G, Ulyanov NB, Camerini-Otero RD, Jernigan RL. 1994. *J. Mol. Biol.* 239:181–200
20. Mirkin SM, Frank-Kamenetskii MD. 1994. *Annu. Rev. Biophys. Biomol. Struct.* 23:541–76
21. Baran N, Lapidot A, Manor H. 1991. *Proc. Natl. Acad. Sci. USA* 88:507–11
22. Samadashwily GM, Dayn A, Mirkin SM. 1993. *EMBO J.* 12:4975–83
23. Rao BS. 1994. *Gene* 140:233–37
24. Hoogsteen K. 1963. *Acta Crystallogr.* 16:907–16
25. Kohwi Y, Kohwi-Shigematsu T. 1988. *Proc. Natl. Acad. Sci. USA* 85:3781–85
26. Beal PA, Dervan PB. 1991. *Science* 251:1360–63
27. Bernues J, Beltran R, Casasnovas JM, Azorin F. 1990. *Nucleic Acids Res.* 18: 4067–73
28. Malkov VA, Voloshin ON, Veselkov AG, Rostapshov VM, Jansen I, et al. 1993. *Nucleic Acids Res.* 21:105–11
29. Malkov VA, Voloshin ON, Soyfer VN, Frank-Kamenetskii MD. 1993. *Nucleic Acids Res.* 21:585–91
30. Potaman VN, Soyfer VN. 1994. *J. Biomol. Struct. Dyn.* 11:1035–40
31. Sklenar V, Feigon J. 1990. *Nature* 345: 836–38

32. Haner R, Dervan PB. 1990. *Biochemistry* 29:9761–5
33. Radhakrishnan I, de los Santos C, Patel DJ. 1991. *J. Mol. Biol.* 221:1403–18
34. Giovannangeli C, Montenay-Garestier T, Rougee M, Chassignol M, Thuong NT, Helene C. 1991. *J. Am. Chem. Soc.* 113:7775–76
35. Roberts RW, Crothers DM. 1991. *Proc. Natl. Acad. Sci. USA* 88:9397–401
36. Kool E. 1991. *J. Am. Chem. Soc.* 113: 6265–66
37. Prakash G, Kool E. 1992. *J. Am. Chem. Soc.* 114:3523–28
38. Booher MA, Wang SH, Kool ET. 1994. *Biochemistry* 33:4645–51
39. Wang SH, Booher MA, Kool ET. 1994. *Biochemistry* 33:4639–44
40. Horne DA, Dervan PB. 1990. *J. Am. Chem. Soc.* 112:2435–37
41. Beal P, Dervan P. 1992. *J. Am. Chem. Soc.* 114:1470–78
42. Jayasena SD, Johnston BH. 1992. *Biochemistry* 31:320–27
43. Jayasena SD, Johnston BH. 1992. *Nucleic Acids Res.* 20:5279–88
44. Jayasena SD, Johnston BH. 1993. *Biochemistry* 32:2800–7
45. Roberts RW, Crothers DM. 1992. *Science* 258:1463–66
46. Escude C, Francois J-C, Sun J-S, Ott G, Sprinzl M, et al. 1993. *Nucleic Acids Res.* 21:5547–53
47. Han H, Dervan PB. 1993. *Proc. Natl. Acad. Sci. USA* 90:3806–10
48. Johnston BH. 1988. *Science* 241:1800–4
49. Voloshin ON, Mirkin SM, Lyamichev VI, Belotserkovskii BP, Frank-Kamenetskii MD. 1988. *Nature* 333:475–76
50. Arnott S, Selsing E. 1974. *J. Mol. Biol.* 88:509–21
51. Macaya R, Schultze P, Feigon J. 1992. *Am. Chem. Soc.* 114:781–83
52. Howard FB, Miles HT, Liu K, Frazier J, Raghunathan G, Sasisekharan V. 1992. *Biochemistry* 31:10671–77
53. Raghunathan G, Miles HT, Sasisekharan V. 1993. *Biochemistry* 32:455–62
54. de los Santos C, Rosen M, Patel D. 1989. *Biochemistry* 28:7282–89
55. Rajagopal P, Feigon J. 1989. *Nature* 339:637–40
56. Mooren MM, Pulleyblank DE, Wijmenga SS, Blommers MJ, Hilbers CW. 1990. *Nucleic Acids Res.* 18: 6523–29
57. Radhukrishnan I, Gao X, de los Santos C, Live D, Patel DJ. 1991. *Biochemistry* 30:9022–30
58. Radhakrishnan I, Patel DJ. 1994. *Structure* 2:17–32
59. Radhakrishnan I, de los Santos C, Patel DJ. 1993. *J. Mol. Biol.* 234:188–97
60. Radhakrishnan I, Patel D. 1993. *Structure* 1:135–52
61. Frank-Kamenetskii MD. 1990. In *DNA Topology and its Biological Effects*, ed. N Cozzarelli, J Wang, pp. 185–215. Cold Spring Harbor: Cold Spring Harbor Lab. Press
62. Lyamichev VI, Mirkin SM, Frank-Kamenetskii MD. 1985. *J. Biomol. Struct. Dyn.* 3:327–38
63. Hentchel CC. 1982. *Nature* 295:714–16
64. Larsen A, Weintraub H. 1982. *Cell* 29: 609–22
65. Mace HAF, Pelham HRB, Travers A. 1983. *Nature* 304:555–57
66. Shen CK. 1983. *Nucleic Acids Res.* 11: 7899–910
67. Cantor CR, Efstratiadis A. 1984. *Nucleic Acids Res.* 12:8059–72
68. McKeon C, Schmidt A, de Crombrugghe BA. 1984. *J. Biol. Chem.* 259:6636–40
69. Christophe D, Cabrer B, Bacolla A, Targovnik H, Pohl V, Vassart G. 1985. *Nucleic Acids Res.* 13:5127–44
70. Margot JB, Hardison RC. 1985. *J. Mol. Biol.* 184:195–210
71. Pulleyblank DE, Haniford DB, Morgan AR. 1985. *Cell* 42:271–80
72. Lilley DM. 1992. *Methods Enzymol.* 212:133–39
73. Hanvey JC, Klysik J, Wells RD. 1988. *J. Biol. Chem.* 263:7386–96
74. Hanvey JC, Shimizu M, Wells RD. 1988. *Proc. Natl. Acad. Sci. USA* 85: 6292–96
75. Htun H, Dahlberg JE. 1988. *Science* 241:1791–96
76. Vojtiskova M, Mirkin S, Lyamichev V, Voloshin O, Frank-Kamenetskii M, Palecek E. 1988. *FEBS Lett.* 234:295–99
77. Htun H, Dahlberg JE. 1989. *Science* 243:1571–76
78. Kang S, Wohlrab F, Wells RD. 1992. *J. Biol. Chem.* 267:1259–64
79. Kang S, Wells RD. 1992. *J. Biol. Chem.* 267:20887–91
80. Shimizu M, Kubo K, Matsumoto U, Shindo H. 1994. *J. Mol. Biol.* 235:185–97
81. Bernues J, Beltran R, Casasnovas JM, Azorin F. 1989. *EMBO J.* 8:2087–94
82. Panyutin IG, Kovalsky OI, Budowsky EI. 1989. *Nucleic Acids Res.* 17:8257–71
83. Dayn A, Samadashwily GM, Mirkin SM. 1992. *Proc. Natl. Acad. Sci. USA* 89:11406–10
84. Kohwi Y. 1989. *Nucleic Acids Res.* 17: 4493–502
85. Beltran R, Martinez-Balbas A, Bernues J, Bowater R, Azorin F. 1993. *J. Mol. Biol.* 230:966–78
86. Kohwi Y, Kohwi-Shigematsu T. 1993. *J. Mol. Biol.* 231:1090–101

87. Martinez-Balbas A, Azorin F. 1993. *Nucleic Acids Res.* 21:2557–62
88. Kohwi-Shigematsu T, Kohwi Y. 1991. *Nucleic Acids Res.* 19:4267–71
89. Panyutin IG, Wells RD. 1992. *J. Biol. Chem.* 267:5495–501
90. Hampel KJ, Ashley C, Lee JS. 1994. *Biochemistry* 33:5674–81
91. Belotserkovskii BP, Veselkov AG, Filippov SA, Dobrynin VN, Mirkin SM, Frank-Kamenetskii MD. 1990. *Nucleic Acids Res.* 18:6621–24
92. Griffin LC, Dervan PB. 1989. *Science* 245:967–71
93. Mergny JL, Sun JS, Rougee M, Montenay-Garestier T, Barcelo F, et al. 1991. *Biochemistry* 30:9791–98
94. Sun JS, Mergny JL, Lavery R, Montenay-Garestier T, Helene C. 1991. *J. Biomol. Struct. Dyn.* 9:411–24
95. Macaya RF, Gilbert DE, Malek S, Sinsheimer JS, Feigon J. 1991. *Science* 254:270–74
96. Strobel SA, Dervan PB. 1990. *Science* 249:73–75
97. Strobel SA, Dervan PB. 1991. *Nature* 350:172–74
98. Strobel SA, Doucette-Stamm LA, Riba L, Housman DE, Dervan PB. 1991. *Science* 254:1639–42
99. Ito T, Smith CL, Cantor CR. 1992. *Proc. Natl. Acad. Sci. USA* 89:495–98
100. Wang E, Malek S, Feigon J. 1992. *Biochemistry* 31:4838–46
101. Kiessling LL, Griffin LC, Dervan PB. 1992. *Biochemistry* 31:2829–34
102. Yoon K, Hobbs CA, Koch J, Sardaro M, Kutny R, Weis AL. 1992. *Proc. Natl. Acad. Sci. USA* 89:3840–44
103. Fossella JA, Kim YJ, Shih H, Richards EG, Fresco JR. 1993. *Nucleic Acids Res.* 21:4511–15
104. Stilz HU, Dervan PB. 1993. *Biochemistry* 32:2177–85
105. Rao TS, Hogan ME, Revankar GR. 1994. *Nucleosides Nucleotides* 13:95–107
106. Beal PA, Dervan PB. 1992. *Nucleic Acids Res.* 20:2773–76
107. Hampel KJ, Crosson P, Lee JS. 1991. *Biochemistry* 30:4455–59
108. Hampel KJ, Burkholder GD, Lee JS. 1993. *Biochemistry* 32:1072–77
109. Tabor CW, Tabor H. 1984. *Annu. Rev. Biochem.* 53:749–91
110. Lee JS, Woodsworth ML, Latimer LJ, Morgan AR. 1984. *Nucleic Acids Res.* 12:6603–14
111. Maher LJ, Wold B, Dervan PB. 1989. *Science* 245:725–30
112. Plum GE, Park YW, Singleton SF, Dervan PB, Breslauer KJ. 1990. *Proc. Natl. Acad. Sci. USA* 87:9436–40
113. Krawczyk SH, Milligan JF, Wadwani S, Moulds C, Froehler BC, Matteucci MD. 1992. *Proc. Natl. Acad. Sci. USA* 89:3761–64
114. Ono A, Ts'o POP, Kan L-S. 1991. *J. Am. Chem. Soc.* 113:4032–33
115. Jetter MC, Hobbs FW. 1993. *Biochemistry* 32:3249–54
116. Koh JS, Dervan PB. 1992. *J. Am. Chem. Soc.* 114:1470–78
117. Sun JS, Giovannangeli C, Francois JC, Kurfurst R, Montenay-Garestier T, et al. 1991. *Proc. Natl. Acad. Sci. USA* 88: 6023–27
118. Mouscadet JF, Ketterle C, Goulaouic H, Carteau S, Subra F, et al. 1994. *Biochemistry* 33:4187–96
119. Orson FM, Kinsey BM, McShan WM. 1994. *Nucleic Acids Res.* 22:479–84
120. Takasugi M, Guendouz A, Chassignol M, Decout JL, Lhomme J, et al. 1991. *Proc. Natl. Acad. Sci. USA* 88:5602–6
121. Mergny JL, Duval-Valentin G, Nguyen CH, Perrouault L, Faucon B, et al. 1992. *Science* 256:1681–84
122. Pilch DS, Waring MJ, Sun JS, Rougee M, Nguyen CH, et al. 1993. *J. Mol. Biol.* 232:926–46
123. Lee JS, Latimer LJ, Hampel KJ. 1993. *Biochemistry* 32:5591–97
124. Belotserkovskii BP, Krasilnikova MM, Veselkov AG, Frank-Kamenetskii MD. 1992. *Nucleic Acids Res.* 20:1903–8
125. Nielsen PE, Egholm M, Berg RH, Buchardt O. 1991. *Science* 254:1497–500
126. Peffer NJ, Hanvey JC, Bisi JE, Thomson SA, Hassman CF, et al. 1993. *Proc. Natl. Acad. Sci. USA* 90:10648–52
127. Egholm M, Buchardt O, Christensen L, Behrens C, Freier SM. 1993. *Nature* 365:566–68
128. Egholm M, Nielsen PE, Buchardt O, Berg RH. 1992. *J. Am. Chem. Soc.* 114: 9677–78
129. Kim SK, Nielsen PE, Egholm M, Buchardt O, Berg RH, Norden B. 1993. *J. Am. Chem. Soc.* 115:6477–81
130. Almarsson O, Bruice TC, Kerr J, Zuckermann RN. 1993. *Proc. Natl. Acad. Sci. USA* 90:7518–22
131. Nielsen PE, Egholm M, Buchardt O. 1994. *J. Mol. Recognit.* 7:165–70
132. Frank-Kamenetskii MD. 1985. In *Structure and Motion: Membranes, Nucleic Acids and Proteins*, ed. E Clementi, G Corongiu, MH Sarma, RH Sarma, et al, pp. 417–43. New York: Adenine
133. Gueron M, Leroy J-L. 1992. Nucleic Acids and Molecular Biology, ed. F Eckstein, DMJ Lilley, pp1–22.Heidelberg: Springer-Verlag
134. Hanvey JC, Peffer NJ, Bisi JE,Thomson

SA, Cadilla R, et al. 1992. *Science* 258: 1481–85
135. Nielsen PE, Egholm M, Berg RH, Buchardt O. 1993. *Anti-Cancer Drug Design* 8:53–63
136. Demidov VV, Potaman VN, Frank-Kamenetskii MD, Egholm M, Buchardt O, Nielsen PE. 1994. *Biochem. Pharmacol.* 48:1310-3
137. Beasty AM, Behe MJ. 1988. *Nucleic Acids Res.* 16:1517–28
138. Manor H, Sridhara-Rao B, Martin RG. 1988. *J. Mol. Evol.* 27:96–101
139. Karlovsky P, Pecinka P, Vojtiskova M, Makaturova E, Palecek E. 1990. *FEBS Lett.* 274:39–42
140. Kohwi Y, Malkhosyan SR, Kohwi-Shigematsu T. 1992. *J. Mol. Biol.* 223: 817–22
141. Ussery DW, Sinden RR. 1993. *Biochemistry* 32:6206–13
142. Wang JC. 1985. *Annu. Rev. Biochem.* 54:665–97
143. Liu LF, Wang JC. 1987. *Proc. Natl. Acad. Sci. USA* 84:7024–27
144. Dayn A, Malkhosyan S, Mirkin SM. 1992. *Nucleic Acids Res.* 20:5991–97
145. Rahmouni AR, Wells RD. 1992. *J. Mol. Biol.* 223:131–44
146. Kohwi Y, Panchenko Y. 1993. *Genes Dev.* 7:1766–78
147. Lee JS, Burkholder GD, Latimer LJ, Haug BL, Braun RP. 1987. *Nucleic Acids Res.* 15:1047–61
148. Burkholder GD, Latimer LJ, Lee JS. 1988. *Chromosoma* 97:185–92
149. Lee JS, Latimer LJ, Haug BL, Pulleyblank DE, Skinner DM, Burkholder GD. 1989. *Gene* 82:191–99
150. Lapidot A, Baran N, Manor H. 1989. *Nucleic Acids Res.* 17:883–900
151. Kornberg A, Baker T. 1992. *DNA Replication.* New York: Freeman. 2nd ed.
152. Rao S, Manor H, Martin RG. 1988. *Nucleic Acids Res.* 16:8077–94
153. Brinton BT, Caddle MS, Heintz NH. 1991. *J. Biol. Chem.* 266:5153–61
154. Gilmour DS, Thomas GH, Elgin SCR. 1989. *Science* 245:1487–90
155. Glaser RL, Thomas GH, Siegfried E, Elgin SC, Lis JT. 1990. *J. Mol. Biol.* 211:751–61
156. Hoffman EK, Trusko SP, Murphy M, George DL. 1990. *Proc. Natl. Acad. Sci. USA* 87:2705–9
157. Lafyatis R, Denhez P, Williams T, Sporn M, Roberts A. 1991. *Nucleic Acids Res.* 19:6419–25
158. Johnson AC, Jinno Y, Merlino GT. 1988. *Mol. Cell Biol.* 8:4174–84
159. Mavrothalassitis GL, Watson DK, Papas TS. 1990. *Proc. Natl. Acad. Sci. USA* 87:1047–51
160. Tewari DS, Cook DM, Taub R. 1989. *J. Biol. Chem.* 264:16238–45
161. Davis TL, Firulli AB, Kinniburgh A. 1989. *Proc. Natl. Acad. Sci. USA* 86: 9682–86
162. Postel EH, Mango SE, Flint SJ. 1989. *Mol. Cell. Biol.* 9:5123–33
163. Clark SP, Lewis CD, Felsenfeld G. 1990. *Nucleic Acids Res.* 18:5119–26
164. Kolluri R, Torrey TA, Kinniburgh AJ. 1992. *Nucleic Acids Res.* 20:111–16
165. Bossone SA, Asselin C, Patel AJ, Marcu KB. 1992. *Proc. Natl. Acad. Sci. USA* 89:7452–56
166. Postel EH, Berberich SJ, Flint SJ, Ferrone CA. 1993. *Science* 261:478–83
167. Brunel F, Alzari PM, Ferrara P, Zakin MM. 1991. *Nucleic Acids Res.* 19: 5237–45
168. Kennedy GC, Rattner JB. 1992. *Proc. Natl. Acad. Sci. USA* 89:11498–502
169. Yee HA, Wong AKC, van de Sande JH, Rattner JB. 1991. *Nucleic Acids Res.* 19:949–53
170. Muraiso T, Nomoto S, Yamazaki H, Mishima Y, Kominari R. 1992. *Nucleic Acids Res.* 20:6631–35
171. Raghu G, Tevosian S, Anant S, Subramanian K, George DL, Mirkin SM. 1994. *Nucleic Acids Res.* 22:3271–79
172. Firulli AB, Maibenko DC, Kinniburgh AJ. 1992. *Biochem. Biophys. Res. Commun.* 185:264–70
173. Firulli AB, Maibenco DC, Kinniburgh AJ. 1994. *Arch. Biochem. Biophys.* 310: 236–42
174. DesJardins E, Hay N. 1993. *Mol. Cell. Biol.* 13:5710–24
175. Kiyama R, Camerini-Otero RD. 1991. *Proc. Natl. Acad. Sci. USA* 88:10450–54
176. Kohwi Y, Kohwi-Shigematsu T. 1991. *Genes Dev.* 5:2547–54
177. Lu Q, Wallrath LL, Allan BD, Glaser RL, Lis JT, Elgin SCR. 1992. *J. Mol. Biol.* 225:985–98
178. Lu Q, Wallrath LL, Granok H, Elgin SCR. 1993. *Mol. Cell. Biol.* 13:2802–14
179. Cooney M, Czernuszewicz G, Postel EH, Flint SJ, Hogan ME. 1988. *Science* 241:456–59
179a. Maher LJ, Oervan PB, Wold B. 1992. *Biochemistry* 31:70–81
180. Gee JE, Blume S, Snyder RC, Ray R, Miller DM. 1992. *J. Biol. Chem.* 267: 11163–67
181. Mayfield C, Ebbinghaus S, Gee J, Jones D, Rodu B, Squibb M, Miller D. 1994. *J. Biol. Chem.* 269:18232–38
182. Grigoriev M, Praseuth D, Robin P, Hemar A, Saison-Behmoaras T, et al. 1992. *J. Biol. Chem.* 267:3389–95

183. Duval-Valentin G, Thuong NT, Helene C. 1992. *Proc. Natl. Acad. Sci. USA* 89:504–8
184. Maher LJ. 1992. *Biochemistry* 31:7587–94
185. Young SL, Krawczyk SH, Matteucci MD, Toole JJ. 1991. *Proc. Natl. Acad. Sci. USA* 88:10023–26
186. Maher LJ, Wold B, Dervan PB. 1991. *Antisense Res. Dev.* 1:277–81
187. Grigoriev M, Praseuth D, Guieysse AL, Robin P, Thuong NT, et al. 1993. *Proc. Natl. Acad. Sci. USA* 90:3501–5
188. Roy C. 1993. *Nucleic Acids Res.* 21: 2845–52
189. Orson FM, Thomas DW, McShan WM, Kessler DJ, Hogan ME. 1991. *Nucleic Acids Res.* 19:3435–41
190. Postel EH, Flint SJ, Kessler DJ, Hogan ME. 1991. *Proc. Natl. Acad. Sci. USA* 88:8227–31
191. McShan WM, Rossen RD, Laughter AH, Trial L, Kessler DJ, et al. 1992. *J. Biol. Chem.* 267:5712–21
192. Ing NH, Beekman JM, Kessler DJ, Murphy M, Jayaraman K, et al. 1993. *Nucleic Acids Res.* 21:2789–96
193. Samadashwily GM, Mirkin SM. 1994. *Gene.* 149:127–36
194. Giovannangeli C, Thuong NT, Helene C. 1993. *Proc. Natl. Acad. Sci. USA* 90:10013–17
195. Hacia JG, Dervan PB, Wold BJ. 1994. *Biochemistry* 33:6192–200
196. Birg F, Praseuth D, Zerial A, Thuong NT, Asseline U, et al. 1990. *Nucleic Acids Res.* 18:2901–8
197. Kowalczykowski SC. 1991. *Annu. Rev. Biochem. Biophys. Chem.* 20: 539–75
198. Stasiak A, Stasiak AZ, Koller T. 1984. *Cold Spring Harbor Symp. Quant. Biol.* 49:561–70
199. Hsieh P, Camerini-Otero CS, Camerini-Otero RD. 1990. *Genes Dev.* 4: 1951–63
200. Camerini-Otero RD, Hsieh P. 1993. *Cell* 73:217–23
201. Wu AM, Kahn R, DasGupta C, Radding CM. 1982. *Cell* 30:37–44
202. Bianchi M, DasGupta C, Radding CM. 1983. *Cell* 34:931–39
203. Umlauf SW, Cox MM, Inman RB. 1990. *J. Biol. Chem.* 265:16898–912
204. Rao BJ, Dutreix M, Radding CM. 1991. *Proc. Natl. Acad. Sci. USA* 88: 2984–88
205. Rao BJ, Chiu SK, Radding CM. 1993. *J. Mol. Biol.* 229:328–43
206. Adzuma K. 1992. *Genes Dev.* 6:1679–94
207. Jain SK, Inman RB, Cox MM. 1992. *J. Biol. Chem.* 267:4215–22
208. Chiu SK, Rao BJ, Story RM, Radding CM. 1993. *Biochemistry* 32:13146–55
209. Shchyolkina AK, Timofeev EN, Borisova OF, Ilicheva IA, Minyat EE, et al. 1994. *FEBS Lett.* 339:113–18
210. Ferrin LJ, Camerini-Otero RD. 1991. *Science* 254:1494–97
211. Bedale WA, Inman RB, Cox MM. 1993. *J. Biol. Chem.* 268:15004–16
212. Hsieh P, Camerini-Otero CS, Camerini-Otero RD. 1992. *Proc. Natl. Acad. Sci. USA* 89:6492–96

1. 5,874,555, Feb. 23, 1999, **Triple helices** and processes for making same; Peter B. Dervan, et al., 536/23.1; 435/6; 436/501; 536/24.1, 24.3, 24.31, 24.32, 24.33, 25.3 [IMAGE AVAILABLE]

2. 5,861,244, Jan. 19, 1999, Genetic sequence assay using **DNA triple strand** formation; Chia-Gee Wang, et al., 435/6; 536/24.3 [IMAGE AVAILABLE]

3. 5,844,110, Dec. 1, 1998, Synthetic **triple helix**-forming compound precursors; Barry I. Gold, 536/29.2, 27.13; 544/279, 283 [IMAGE AVAILABLE]

4. 5,834,185, Nov. 10, 1998, Formation of **triple helix** complexes of single **stranded nucleic acids** using nucleoside oligomers which comprise pyrimidine analogs, **triple helix** complexes formed thereby and oligomers used in their formation; Paul On-Pong Ts'o, et al., 435/6; 436/501; 514/44; 536/25.3 [IMAGE AVAILABLE]

5. 5,800,984, Sep. 1, 1998, **Nucleic acid** sequence detection by **triple helix** formation at primer site in amplification reactions; Calvin P. H. Vary, 435/6, 91.1, 91.2; 536/24.3, 24.31, 24.32, 24.33 [IMAGE AVAILABLE]

6. 5,789,155, Aug. 4, 1998, Process for identifying **nucleic acids** and **triple helices** formed thereby; Peter B. Dervan, et al., 435/6; 436/501; 536/23.1, 24.1, 24.3, 24.31, 24.32, 24.33, 25.3 [IMAGE AVAILABLE]

7. 5,693,471, Dec. 2, 1997, **Triple stranded nucleic acids**; Jacques R. Fresco, 435/6, 5, 91.1; 514/44; 536/24.5 [IMAGE AVAILABLE]

8. 5,683,874, Nov. 4, 1997, Single-**stranded** circular oligonucleotides capable of forming a **triplex** with a target sequence; Eric T. Kool, 435/6; 514/44; 536/23.1, 24.3, 24.5 [IMAGE AVAILABLE]

9. 5,665,541, Sep. 9, 1997, Formation of **triple helix** complexes for the detection of double **stranded DNA** sequences using oligomers which comprise an 8-modified purine base; Paul S. Miller, et al., 435/6; 536/23.1, 24.3, 25.3 [IMAGE AVAILABLE]

10. 5,645,985, Jul. 8, 1997, Enhanced **triple-helix** and double-**helix** formation with oligomers containing modified pyrimidines; Brian Froehler, et al., 435/6; 536/24.3, 24.31, 24.32, 24.5, 26.8 [IMAGE AVAILABLE]

11. 5,594,121, Jan. 14, 1997, Enhanced **triple-helix** and double-**helix** formation with oligomers containing modified purines; Brian Froehler, et al., 536/23.5; 435/6, 91.1; 536/23.1, 24.3, 24.5, 27.2, 27.6, 27.81 [IMAGE AVAILABLE]

12. 5,422,251, Jun. 6, 1995, **Triple-stranded nucleic acids**; Jacques R. Fresco, 435/91.1, 6; 536/24.31, 24.32 [IMAGE AVAILABLE]

13. 5,401,632, Mar. 28, 1995, **Triple helix** purification and sequencing; Renfeng Wang, et al., 435/6, 91.5; 536/23.1, 24.3, 24.33 [IMAGE AVAILABLE]

12

STIC-ILL

From: Enewold, Jeanine
Sent: Thursday, July 08, 1999 8:44 AM
To: STIC-ILL
Subject: Art- 09137822

Hi-
Could you please find and copy these articles.

1. Bioorg Med Chem 1996 Jan;4(1):5-23
"Peptide nucleic acids (PNA): synthesis, properties and potential applications."
Hyrup B, Nielsen PE

2. Science 1995 Dec 15;270(5243):1838-41
"A nucleic acid triple helix formed by a peptide nucleic acid-DNA complex."
Betts L, Josey JA, Veal JM, Jordan SR

3. Annu Rev Biochem 1995;64:65-95
"Triplex DNA structures."
Frank-Kamenetskii MD, Mirkin SM

4. J Biol Chem 1995 Jun 9;270(23):14068-71
"An unusually stable purine(purine-pyrimidine) short triplex. The third strand stabilizes double-stranded DNA."
Svinarchuk F, Paoletti J, Malvy C

Thank you

Jeanine Enewold
Art Unit: 1655
Room 12-D11
306-5817

tion in the dye nuclear/bath fluorescence = 0.15 ± 0.09 ... 5). The addition of $InsP_3$ to extruded nucleoplasm did not change the calcium green fluorescence (n = 3), indicating that the lack of fluorescence from the nucleus after $InsP_3$ application was not caused by a direct interaction between components of the nucleoplasm and $InsP_3$. These results eliminate the possibility that the nuclear matrix was necessary for dye exclusion and suggest that the envelope itself regulates transit of intermediate-sized (10 kD) molecules.

Although intermediate-sized molecules of ~10 kD were excluded by the nuclear envelope when the nuclear cisterna Ca^{2+} store was emptied, smaller molecules were not. Intact nuclei were Ca^{2+}-depleted by incubation in 10 nM Ca^{2+} solution with 1 μM $InsP_3$ for 3 min. Nuclei were exposed to a low molecular weight form of Lucifer yellow (500 daltons). Within 2 min after the addition of Lucifer yellow, nuclei had roughly the same fluorescence intensity as the bath (nuclear/bath = 0.84 ± 0.04, n = 8). To determine whether ions were excluded from nuclei, we loaded intact nuclei with the salt form of indo-1 and the nuclear Ca^{2+} store depleted with $InsP_3$. After depletion of stores, 100 μM Mn^{2+} was added to the bath and the fluorescence within the nucleoplasm was monitored. All fluorescence in the nucleoplasm was quenched with the addition of Mn^{2+}, without regard to the amount of Ca^{2+} within the store. In fact, the rate of quench and final fluorescence values were not statistically different in conditions of Ca^{2+}-filled (Fig. 1C) or Ca^{2+}-depleted nuclear stores (n = 8) (Fig. 4C). Lower $[Mn^{2+}]$ (<1 μM) had similar effects (n = 4). Taken together, the results suggest that the nuclear Ca^{2+} store regulates the movement of molecules of ~10 kD, but not smaller molecules or ions.

Our results demonstrate that the nuclear Ca^{2+} store regulates size-specific entry of molecules into the *Xenopus* oocyte nucleus, including intermediate-sized molecules (10 kD) lacking an NLS that were previously thought to pass freely through the nuclear pore complex. We demonstrated that (i) movement of 10-kD molecules across the nuclear envelope depended on Ca^{2+} within the nuclear cisterna; (ii) depletion of this store was sufficient to halt diffusion across the envelope and did not require the nuclear matrix; and (iii) molecules and ions <500 daltons crossed the nuclear envelope regardless of the state of the nuclear Ca^{2+} store. The mechanism by which store depletion is sensed by the nuclear pore is unknown, but the nuclear pore protein, gp210, contains multiple Ca^{2+}-binding domains predicted to reside within the nuclear cisternae (*11*). Such a molecule might sense depletion of the nuclear cisternal $[Ca^{2+}]$ and initiate conformational changes that block intermediate-sized molecule diffusion into the nucleus.

REFERENCES AND NOTES

1. C. Dingwall and R. Laskey, *Science* **258**, 942 (1992); C. Dingwall, *BioEssays* **13**, 213 (1991).
2. L. Davis, *Curr. Opin. Cell Biol.* **4**, 424 (1992).
3. C. W. Akey, *J. Cell Biol.* **109**, 955 (1989).
4. E. C. Hurt, *FEBS Lett.* **325**, 76 (1993).
5. M. S. Moore and G. Blobel, *Cell* **69**, 939 (1992); M. Stewart, *Cell Biol.* **3**, 267 (1992); M. A. Powers and D. J. Forbes, *Cell* **79**, 931 (1994).
6. I. Lang, M. Scholz, R. Peters, *J. Cell Biol.* **102**, 1183 (1986); L. Gerace and B. Burke, *Annu. Rev. Cell Biol.* **4**, 335 (1988); P. A. Silver, *Cell* **64**, 489 (1991).
7. L. Stehno-Bittel, A. Luckhoff, D. E. Clapham, *Neuron* **14**, 163 (1995).
8. D. O. Mak and J. K. Foskett, *J. Biol. Chem.* **269**, 29375 (1994).
9. P. Nicotera, S. Orrenius, T. Nilsson, P. O. Berggren, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6868 (1990); A. N. Malviya, P. Rogue, G. Vincendon, *ibid.*, p. 9270.
10. G. Grynkiewicz, M. Poenie, R. Y. Tsien, *J. Biol. Chem.* **260**, 3440 (1985); T. J. Hallam and T. J. Rink, *FEBS Lett.* **188**, 175 (1985); C. Lin, G. Hajnoczky, A. P. Thomas, *Cell Calcium* **16**, 247 (1994).
11. U. F. Greber and L. Gerace, *J. Cell Biol.* **128**, 5 (1995).
12. O. Bachs and N. C. E. Agell, *Biochim. Biophys. Acta* **1113**, 259 (1992).
13. *Xenopus laevis* oocytes were removed from toads as described [J. D. Lechleiter and D. E. Clapham, *Cell* **69**, 283 (1992)]. The salt form of fluorescent dyes was injected (50 nl of 1.0 mM dye/oocyte) into oocytes. After 30 min, defolliculated oocytes were enucleated manually by bisecting the oocyte along its equator. Nuclei were removed and washed of cytoplasm in mock intracellular solution containing 140 mM KCl, 10 mM Hepes, 3 mM $MgCl_2$ (pH 7.2). Loading of the nuclear cisterna with membrane-permeant fluorescent dyes was accomplished by incubating the isolated nucleus in mock intracellular solution containing the 100 μM AM form of dye for 15 to 30 min. All fluorescent molecules were purchased from Molecular Probes (Eugene, OR). Nuclei were imaged on a Bio-Rad confocal microscope as described (*8*) or on a Zeiss LSM-410 Laser Scanning confocal microscope. The light source was an argon/krypton laser tuned to either 360 or 488 nm. Two-dimensional confocal images were acquired by scanning 256 by 256 pixels per image. The effective thickness of the confocal optical section was controlled by varying the opening of a pinhole aperture in the detection path. Images were taken every 2 to 5 s, depending on the protocol, with a gray scale value for fluorescence intensity from 1 to 255. Images were processed on a Silicon Graphics Iris computer with ANALYZE software (Mayo Foundation). Background fluorescence values were subtracted in pseudocolor images (Figs. 1 and 3); black represents 0 and blue represents ~1 to 35 arbitrary pixel intensity units. Pixel values were averaged from within the nucleus and divided by the average pixel value of the bath. A ratio value of 1 indicated that fluorescence was equivalent in the nucleus and the bath. Results were expressed as means ± SEM. All experiments were done at room temperature (22° ± 2°C).
14. Nuclear ghosts were formed most commonly by positioning a small air bubble at the end of a micropipette and placing it in contact with the surface of isolated nuclei. Other methods included manually removing the nucleoplasm with a glass micropipette used for microinjection of mammalian cells. Typically, the nucleoplasm required 10 to 15 min to completely exit the nuclear envelope.
15. G. Schlenstedt, E. Hurt, V. Doye, P. A. Silver, *J. Cell Biol.* **123**, 785 (1993).
16. P. L. Paine, M. E. Johnson, Y. T. Lau, L. J. M. Tluczek, D. S. Miller, *BioTechniques* **13**, 238 (1992).
17. M. S. Schmidt-Zachmann, C. Dargemont, L. C. Kuhn, E. A. Nigg, *Cell* **74**, 493 (1993).

18. Supported by NIH grant 41303 to D.E.C.

13 July 1995; accepted 18 October 1995

A Nucleic Acid Triple Helix Formed by a Peptide Nucleic Acid–DNA Complex

Laurie Betts, John A. Josey,* James M. Veal, Steven R. Jordan

The crystal structure of a nucleic acid triplex reveals a helix, designated P-form, that differs from previously reported nucleic acid structures. The triplex consists of one polypurine DNA strand complexed to a polypyrimidine hairpin peptide nucleic acid (PNA) and was successfully designed to promote Watson-Crick and Hoogsteen base pairing. The P-form helix is underwound, with a base tilt similar to B-form DNA. The bases are displaced from the helix axis even more than in A-form DNA. Hydrogen bonds between the DNA backbone and the Hoogsteen PNA backbone explain the observation that polypyrimidine PNA sequences form highly stable 2:1 PNA-DNA complexes. This structure expands the number of known stable helical forms that nucleic acids can adopt.

Oligonucleotides and modified derivatives, such as phosphorothioates, are being examined as antisense therapeutic agents. An ideal antisense agent would be nuclease-resistant and uncharged to aid cellular penetration. A dramatic deviation from the phosphoribose backbone is PNA, originally developed to be an antigene triplexing agent (*1*). PNA is uncharged and stable toward nucleases, and the achiral PNA backbone (Fig. 1A) can be synthesized by amide-based chemistry.

The first report on PNA demonstrated that a T_{10} PNA polymer bound its complementary A_{10} DNA sequence with a markedly increased T_m (temperature at which 50% of double-stranded DNA is

Genetics Institute, 87 Cambridge Park Drive, Box 2, Cambridge, MA 02140, USA

*Present address: Amgen Boulder, 3200 Walnut Street, Boulder, CO 80301, USA

...tive to a T_{10}·A_{10} DNA duplex (1). ... T_{10} PNA displaced the polydhymidylate [poly(T)] portion of a DNA duplex target (*1*, *2*) at low ionic strength. Further study showed a tendency for poly(T) PNA to form 2:1 PNA-DNA complexes (*3*). Several groups have designed and synthesized *bis*- or hairpin PNAs to promote triplex formation by tethering two polypyrimidine PNA strands by flexible linkers (*4*, *5*). These PNAs did indeed have increased affinity for single-stranded DNA and a higher rate of strand invasion for double-stranded DNA, relative to single-stranded PNA. To gain insight into the manner in which PNAs complex with DNA to form triplexes, we determined the structure of a nine-base hairpin PNA-DNA complex (Fig. 1B). The structure provides a basis for understanding the high affinity toward nucleic acids exhibited by polypyrimidine PNA and reveals an unusual helix.

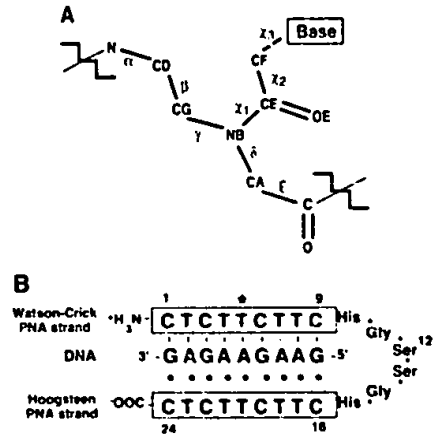

Fig. 1. PNA monomer and PNA_2-DNA_2 complex structures. (**A**) Structure of a PNA monomer. Backbone torsion angles are indicated by Greek letters according to convention (*11*). Carbon positions are designated by A, D, G, E, and F; nitrogen position by B; and oxygen position by E. (**B**) Diagram of the PNA-DNA triplex. DNA was either synthesized by us or purchased from Research Genetics and used without further purification. PNAs were synthesized as described (*4*). The 5-iodo-U base (*T) provided phase information and was a convenient reference point in the electron density. The hexapeptide linker, His-Gly-Ser-Ser-Gly-His, consists of all (L) amino acids.

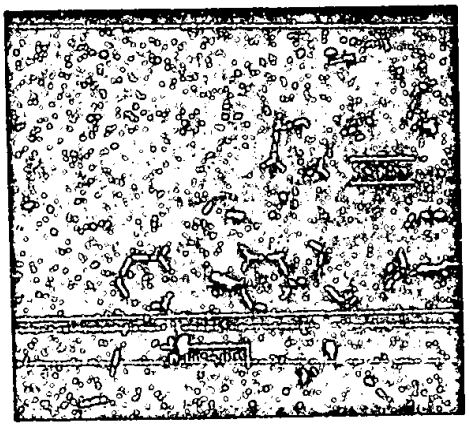

Fig. 2. Representative C·G·C base triplet and its corresponding electron density. The map was calculated with the 2.5 Å symmetry-averaged, solvent-flattened SIRAS phases and is contoured at 1σ.

Table 1. Structure determination. Hairpin PNA-DNA complexes were prepared for crystallization by annealing. Complex at 0.3 mM was mixed with an equal volume of 1.4 M ammonium sulfate and 0.1 M tris-HCl (pH 8.5) and equilibrated at 22°C or 4°C (hanging drop or dialysis buttons) against a reservoir at 1.4 M ammonium sulfate and 0.1 M tris. Crystals were space group $P6_522$, $a = b = 73.38$, $c = 141.28$ Å. Native data were collected at −170°C at the CHESS A-1 beamline to 2.8 Å, with a crystal equilibrated in mother liquor including 25% glycerol. Data were reduced and scaled with DENZO (*16*). 5-Iodo-U derivative crystal data were obtained to a resolution of 2.5 Å at room temperature on a rotating anode with *R*-axis II imaging plates and processed with *R*-axis software. The two iodine atoms in the asymmetric unit were located by difference Patterson maps and confirmed by anomalous difference Pattersons. Initial SIRAS phases to 4.0 Å were calculated and refined with the PHASES program (*17*). Solvent flattening was done to 4.0 Å until convergence. Heavy atom parameters were refined against solvent-flattened phases, iterating the process in increasing shells of resolution to 2.8 Å. A starting model was then built into the electron density with the O graphics program (*18*). The crude model consisting of the bases and backbone provided a mask for noncrystallographic symmetry averaging and phase extension to 2.5 Å. The model was rebuilt into the 2.5 Å map and had a crystallographic *R* factor of 45%. X-PLOR simulated annealing refinement (*19*) with AMBER-based (*20*) parameters with the iodo-U crystal data gave a model with an *R* factor of 22% from 6.0 to 2.5 Å. Thirty-five solvent molecules were identified in $F_o - F_c$ maps by using a 2.8σ cutoff. The final model after individual *B* factor refinement gave an *R* factor of 18.7%. FOM, figure of merit; rms, root mean square.

The structure was solved by using a PNA with an iodinated base to provide initial phase information (Table 1). The SIRAS phases were refined by solvent flattening and noncrystallographic symmetry averaging of the two triplexes in the asymmetric unit. The resulting 2.5 Å electron density map (Fig. 2) clearly shows the positions of

Data collection		
Parameter	Native	5-Iodo-U PNA
Resolution (Å)	2.8	2.5
Total observations	52,364	29,543
Unique reflections	5,723	8,374
Completeness (%)	95	93
R_{sym}* (%)	4.1	4.8
SIRAS statistics		
Statistic	Isomorphous	Anomalous
Phasing power	1.41	3.31
R_{Cullis}†	0.580	
Reflections used (*n*)	5,355	4,149‡
Overall FOM	0.607	
Refinement		
Resolution (Å)	6.0–2.5	
R factor (R_{free})§ (%)	18.7 (23.2)	
Average I/σ to 2.5 Å	6.2	
Reflections with $\|F\| > 2\sigma$	6,857	
Total number of atoms	1,941	
Water molecules (*n*)	35	
rms deviation of bond lengths (Å)	0.014	
rms deviation of bond angles (degrees)	2.43	

*R_{sym} is the agreement between all observations of symmetry-related reflections. †R_{Cullis} (centric) = $\Sigma\|\|F_{PH(obs)}\| \pm |F_{P(obs)}\| - |F_{PH(calc)}\|\|/\Sigma\|F_{PH(obs)}| \pm |F_{P(obs)}\|$, where $F_{PH(obs)}$ and $F_{P(obs)}$ are the observed derivative and native structure amplitudes, respectively, and $F_{PH(calc)}$ is the calculated heavy atom structure factor. ‡Number of reflections with average anomalous difference >2σ; anomalous difference = $|I^+ - I^-|$, where *I* is intensity. §R_{free} is the cross-validation *R* factor computed for the test set of reflections (8% of the total), which are omitted in the refinement process.

Table 2. Average torsion angles and helical parameters of PNA-DNA triplex compared with canonical A-DNA and B-DNA. The average angle for each torsion was calculated over both triplexes in the asymmetric unit. Helical parameters for the DNA portion of the triplex were determined with programs described in (*21*).

Molecule	Torsion angles (degrees)									
	α	β	γ	δ	ε	ξ	χ	χ_1	χ_2	χ_3
DNA in triplex	−70	173	61	77	−161	−69	−167			
A-DNA	−50	172	41	79	−146	−78	−154			
B-DNA*	−46	−147	36	157	155	−96	−98			
PNA in WC strand	−103	73	70	93	165			1	−170	89
PNA in H strand	−108	69	69	87	175			1	−175	102

Helical parameters					
	Twist (degrees)	Rise (Å)	Base tilt (Å)	Displacement (Å)	Bases per turn
A-DNA	32.7	2.6	20.0	4.5	11
B-DNA	36.0	3.4	−5.9	−0.1	10
DNA in triplex	22.9	3.4	5.1	6.8	16

*From table 9.3, pp. 230–232, in (*9*).

all the nucleic acid atoms and all of the peptide linker backbone.

The structure revealed a helix (Fig. 3A), which we refer to as P-form for PNA, with helical parameters distinct from either A-form or B-form DNA (Table 2). The P-form helix has a large cavity along the helix axis, reflected by an average base displacement of 6.8 Å (Fig. 3B), compared with 4.5 Å for A-form DNA. The deoxyribose sugars all have a C3′-endo conformation with an average interphosphate distance of 6.0 Å, similar to A-form DNA. This conformation is consistent with the fact that PNAs, including hairpins, bind more tightly to RNA than DNA (5). The tilt of the base triplets, however, is more similar to that of B-form DNA, where the bases are nearly perpendicular to the helix axis. The relative orientation of the Watson-Crick (WC) strands is such that the NH_2-terminus of the PNA is aligned with the 3′ end of the DNA strand. This strand orientation preference is consistent with studies of mixed base sequences and hairpin polypyrimidine PNAs (4–6). The Hoogsteen (H) strand of the PNA is antiparallel to the WC PNA strand, as designed.

Within the two triplexes in the asymmetric unit all 10 T·A–T triplets and 7 of the 8 C·G-C triplets form with the expected hydrogen-bonding geometry and distances. The purine base of each triplet forms a Watson-Crick base pair with the pyrimidine base on one PNA strand and a Hoogsteen base pair with the pyrimidine on the other strand. This type of (Y·R-Y) interaction has been seen in nuclear magnetic resonance (NMR) studies of oligonucleotide triplexes (7) and predicted for PNA_2-DNA_1 complexes (3–6). The average distance between the N3 atoms of the Hoogsteen cytosines and N7 atoms of the guanines is 2.8 Å, indicating that those atoms must be hydrogen bonded (Fig. 2). The N3 of the Hoogsteen cytosines must be significantly protonated even though the pK_a (where K_a is the acidity constant) of a free cytosine is 4.2 and this triplex was crystallized at pH 8.5. The one cytosine that does not make the intended Hoogsteen interaction, cytosine[16], swings out to form an edge to face interaction with cytosine[6] of a crystallographically related complex.

The pattern of base stacking along the Watson-Crick portion of the P-form triplex resembles that found in an A-form DNA duplex (8), despite the much larger displacement of the bases from the P-form helix axis. The 2-keto groups of the Hoogsteen pyrimidines stack over the imidazole ring of the preceding DNA purine. These interactions are consistent with those in

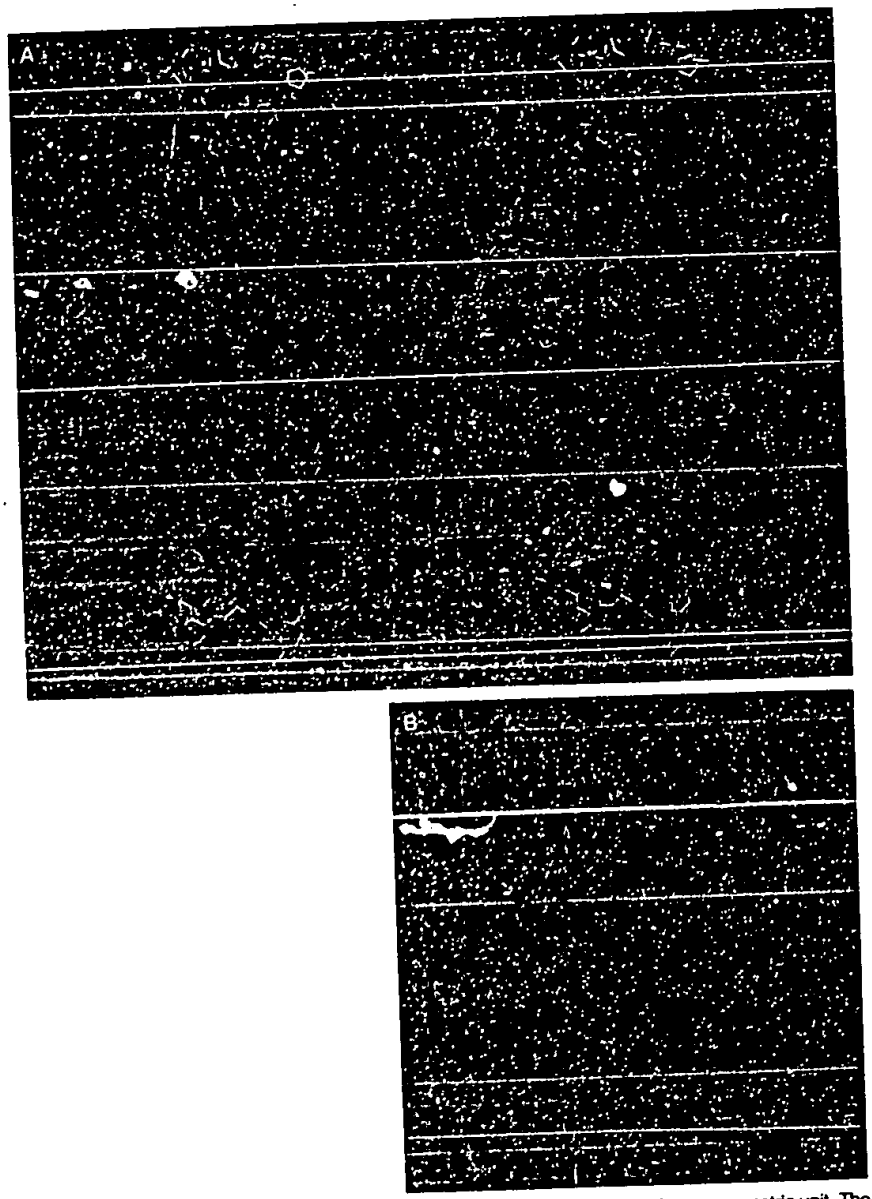

Fig. 3. Structure of the PNA_2-DNA_1 triplex. **(A)** Stereo view of the two triplexes in an asymmetric unit. The DNA in both triplexes is yellow. The magenta strands are the H PNA strands, and the cyan strands are WC PNA strands. Hydrogen bonds between the H PNA and DNA backbone are indicated by white lines. Linker amino acids are drawn in white. **(B)** Comparison of B-, P-, and A-form helices. The H strand of the P-form triplex is magenta. The diameter of the P-form is ~26 Å, whereas that of the A-form is 20 Å. The A- and B-DNA models were built by the program QUANTA (15) with the same DNA sequence as the P-form, with each model representing slightly more than one helical turn.

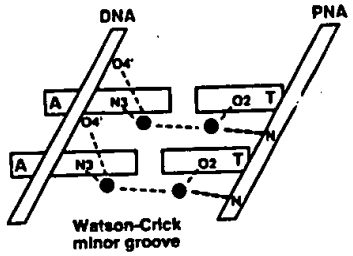

Fig. 4. Minor groove interactions in an A-T region. Solvent molecules are black spheres. Hydrogen bonds are drawn as dashed lines. The DNA backbone is filled in gray, and the PNA is white. A water molecule links each WC PNA backbone amide proton to the O2 keto oxygen of the preceding pyrimidine residue on that strand. A second row of waters bridges the first row with the ribose O4′ atom and adenine N3 atoms. As a consequence of this hydration network, the WC PNA and DNA backbones are linked together.

which polar atoms are stacked over the aromatic rings of adjacent bases (9).

The triplex structure revealed a strong interaction between the DNA backbone and the PNA backbone of the Hoogsteen strand. The O1P oxygen from each phosphate group of the DNA backbone hydrogen bonds to the amide proton of each residue of the PNA backbone of the H strand with an average distance of 2.85 Å. There are extensive van der Waals interactions between the two strands. A total of 1180 Å^2 out of 260 Å^2 of surface area per H PNA residue is buried on complex formation with a DNA residue. In contrast, the WC PNA backbone makes no direct interactions with the DNA.

There are two triplexes stacked in the asymmetric unit, related by approximate twofold symmetry perpendicular to the helical axis (Fig. 3A). As a result, the two PNA backbones form a nearly continuous helix with the carboxylate of the COOH-terminus of each of the H PNA chains forming a salt bridge with the NH_3-terminus of each WC PNA chain. The amino acid linkers sterically block any potential stacking between the top of each triplex and any other triplex, preventing formation of an infinite helix.

The linker peptide was chosen to be both flexible and to impart solubility. Each linker in the asymmetric unit adopts a different conformation although all phi-psi angles are within sterically allowed regions. The average main-chain temperature factor for the amino acid linker region is higher than that of the PNA backbone, further indicating that the linker is relatively flexible and does not impose a particular conformation on the triple helix.

There are 35 clearly defined water molecules. One class of water molecules binds in the minor groove to both the amide hydrogen of the WC PNA backbone, with an average distance of 2.9 Å, and to the O2 oxygen of the preceding pyrimidine base, with an average distance of 3.0 Å (Fig. 4). These water molecules stabilize the residues in the WC PNA strand in a conformation that is nearly identical to the H PNA strand. They were the strongest peaks in the $F_o - F_c$ difference maps (where F_o and F_c are the observed and calculated structure factors, respectively) and were present in the original electron density. Another set of water molecules bridges the first set to the N3 atoms of the adenine bases in the minor groove (Fig. 4). In the case of C·G-C triplets, there is no corresponding water, presumably because of steric interference by the exocyclic amino group of the guanine. These waters all interact to stabilize the P-form helix. Minor groove solvent networks have been observed to stabilize B-form DNA in A-T rich regions by the so-called spine of hydration (*10*). In the P-form major groove, exocyclic heteroatoms of the pyrimidines are generally solvated with either one or two water molecules.

The structure of the complex explains a number of biochemical and biophysical properties of PNAs. The structure provides direct confirmation that PNAs can recognize nucleic acids by both Watson-Crick and Hoogsteen hydrogen bonding. The conformation of the PNA and DNA backbones that accomodates both duplex and triplex interactions is elucidated by the structure (Table 2). The tight association between the DNA strand and the H PNA strand is stabilized both by hydrogen bonds between the DNA phosphate oxygens and the PNA amides and by extensive van der Waals interactions that lock the structure in the P-form helical conformation. Hydrogen bonds to solvent molecules in the minor groove strengthen the P-form Watson-Crick PNA-DNA interactions. This stable triplex accounts for the strong tendency of polypyrimidine PNA sequences to associate with polypurine DNA in a 2:1 ratio and also accounts for the increased T_m of these complexes compared with DNA duplex and triplex sequences.

The NMR structure of a mixed sequence PNA-RNA duplex has been recently reported (*11*) and was described to be most consistent with an A-form helix. The PNA χ_1 torsion angle is very similar in both the NMR duplex and x-ray triplex structures, so that the carbonyl of the tertiary amide points toward the COOH-terminus of the PNA strand. The duplex and triplex structures differ at the α and ε torsion angles, which specify the orientation of the interresidue amide carbonyl. In the refined NMR duplex structures, the predominant orientation of the backbone carbonyl oxygen is more toward the PNA NH_3-terminus, whereas in the P-form triplex the carbonyl oxygen points toward the PNA COOH-terminus. In the triplex crystal structure, the conformation of the PNA residues is nearly identical in the WC strand and the H strand, suggesting that it is a stable conformation consistent with forming either a WC duplex or a triplex. The observed differences in backbone conformation between the NMR duplex and the x-ray triplex could arise from factors such as differences in sequence length and composition, solution compared with crystal, and an RNA compared with a DNA target.

DNA triplex structures have been studied predominantly by NMR. In some studies the triplex is described as being more similar to B-form (*7*), whereas other studies indicate it is more similar to A-form (*12*), as reported for the original fiber diffraction studies of poly(dT)·poly(dA)·poly(dT) (*13*). Base triplets have been observed in a number of crystal structures, but the extent of the "helix" is limited to one or two base triplets, usually resulting from an interaction between an overhanging base with a duplex strand (*14*). Clearly, triplex formation may require features that are not easily accommodated by either classical A- or B-form helices. The PNA_2-DNA_1 triplex forms a previously unknown helix that expands the library of known nucleotide helical structures. This study suggests that there may be additional classes of stable DNA triplexes yet to be discovered.

REFERENCES AND NOTES

1. P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **254**, 1497 (1991).
2. D. Y. Cherny *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1667 (1993); N. J. Peffer *et al.*, *ibid.*, p. 10648.
3. M. Egholm, O. Buchardt, P. E. Nielsen, R. Berg, *J. Am. Chem. Soc.* **114**, 1895 (1992); M. Egholm, P. E. Nielsen, O. Buchardt, R. H. Berg, *ibid.*, p. 9677; S. K. Kim *et al.*, *ibid.* **115**, 6477 (1992).
4. J. A. Josey *et al.*, in preparation.
5. M. Egholm *et al.*, *Nature* **365**, 566 (1993); M. C. Griffith *et al.*, *J. Am. Chem. Soc.* **117**, 831 (1995); M. Egholm *et al.*, *Nucleic Acids Res.* **23**, 217 (1995).
6. P. E. Nielsen, M. Egholm, O. Buchardt, *J. Mol. Recognit.* **7**, 165 (1995).
7. R. Macaya, E. Wang, P. Schultze, V. Sklenar, J. Feigon, *J. Mol. Biol.* **225**, 755 (1992); I. Radhakrishnan and D. J. Patel, *Structure* **2**, 17 (1994).
8. W. N. Hunter, B. Langlois-D'Estaintot, O. Kennard, *Biochemistry* **28**, 2444 (1989).
9. W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984), pp. 132–133.
10. H. R. Drew and R. E. Dickerson, *J. Mol. Biol.* **151**, 535 (1981).
11. S. C. Brown, S. A. Thomson, J. M. Veal, D. G. Davis, *Science* **265**, 777 (1994).
12. C. M. de los Santos, D. Rosen, D. J. Patel, *Biochemistry* **28**, 7282 (1989); K. Umemoto, M. H. Sarma, G. Gupta, J. Luo, R. H. Sarma, *J. Am. Chem. Soc.* **112**, 4539 (1990).
13. S. Arnott and E. Selsing, *J. Mol. Biol.* **88**, 509 (1974).
14. L. Van Meervelt *et al.*, *Nature* **374**, 742 (1995).
15. QUANTA, 4.0; developed by Molecular Simulations, Burlington, MA (1994).
16. Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, Daresbury, England, 29 and 30 January 1993; L. Sawyer, N. Isaacs, S. Bailey, Eds. (Science and Engineering Research Council Daresbury Laboratory, England, 1993), pp. 56–62.
17. W. Furey and S. Swaminathan, *Am. Crystallogr. Assoc. Program Abstr.* **18**, 73 (1990).
18. A. T. Jones and M. Kjeldgaard, O manual, version 5.9.1 (University of Uppsala, Uppsala, Sweden, 1994).
19. A. T. Brunger, X-PLOR manual, version 3.1 (Yale Univ. Press, New Haven, 1987).
20. S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, *J. Comput. Chem.* **7**, 230 (1986).
21. M. S. Babcock, E. P. D. Pednault, W. K. Olson, *J. Biomol. Struct. Dyn.* **11**, 597 (1993).
22. We thank S. C. Brown for discussions, F. Hassman for synthesis of DNA, J-M. Ding and D. Rose for analysis of dissolved crystals, and S. Ginell and L. Keefe (Brookhaven National Laboratory) for help in data collection on a brominated crystal. The Brookhaven Protein Data Bank accession code for the PNA-DNA triplex is 1PNN.

7 June 1995; accepted 18 October 1995

STIC-ILL

From: Enewold, Jeanine
Sent: Thursday, July 08, 1999 12:45 PM
To: STIC-ILL
Subject: Art --09/137822 (B)

12

Please copy these articles....

Trends Biotechnol 1997 Jun;15(6):224-9
"Peptide nucleic acids: expanding the scope of nucleic acid recognition."
Corey DR

Q Rev Biophys 1996 Dec;29(4):369-94
"PNA-nucleic acid complexes. Structure, stability and dynamics."
Eriksson M, Nielsen PE

Trends Biotechnol 1993 Sep;11(9):384-6
"Peptide nucleic acids and their potential applications in biotechnology."
Buchardt O, Egholm M, Berg RH, Nielsen PE

THANK YOU
Jeanine Enewold
Art Unit: 1655
Room 12-D11
306-5817